Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

Generalized Fiducial Inference: A Review and New Results

Jan Hannig, Hari Iyer, Randy C. S. Lai & Thomas C. M. Lee

To cite this article: Jan Hannig, Hari Iyer, Randy C. S. Lai & Thomas C. M. Lee (2016) Generalized Fiducial Inference: A Review and New Results, Journal of the American Statistical Association, 111:515, 1346-1361, DOI: 10.1080/01621459.2016.1165102

To link to this article: https://doi.org/10.1080/01621459.2016.1165102

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2016, VOL. 111, NO. 515, 1346–1361, Review
http://dx.doi.org/10.1080/01621459.2016.1165102

Taylor & Francis
Taylor & Francis Group

REVIEW

Generalized Fiducial Inference: A Review and New Results

Jan Hannig[a], Hari Iyer[b], Randy C. S. Lai[c,d], and Thomas C. M. Lee[c]

[a]Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC, USA; [b]Statistical Engineering Division, National Institute of Standards and Technology, Gaithersburg, MD, USA; [c]Department of Statistics, University of California, Davis, Davis, CA, USA; [d]Department of Mathematics & Statistics, University of Maine, Orono, ME, USA

ABSTRACT

R. A. Fisher, the father of modern statistics, proposed the idea of fiducial inference during the first half of the 20th century. While his proposal led to interesting methods for quantifying uncertainty, other prominent statisticians of the time did not accept Fisher's approach as it became apparent that some of Fisher's bold claims about the properties of fiducial distribution did not hold up for multi-parameter problems. Beginning around the year 2000, the authors and collaborators started to reinvestigate the idea of fiducial inference and discovered that Fisher's approach, when properly generalized, would open doors to solve many important and difficult inference problems. They termed their generalization of Fisher's idea as generalized fiducial inference (GFI). The main idea of GFI is to carefully transfer randomness from the data to the parameter space using an inverse of a data-generating equation without the use of Bayes' theorem. The resulting generalized fiducial distribution (GFD) can then be used for inference. After more than a decade of investigations, the authors and collaborators have developed a unifying theory for GFI, and provided GFI solutions to many challenging practical problems in different fields of science and industry. Overall, they have demonstrated that GFI is a valid, useful, and promising approach for conducting statistical inference. The goal of this article is to deliver a timely and concise introduction to GFI, to present some of the latest results, as well as to list some related open research problems. It is authors' hope that their contributions to GFI will stimulate the growth and usage of this exciting approach for statistical inference. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received February 2015
Revised January 2016

KEYWORDS
Approximate Bayesian computations; Data-generating equation; Fiducial inference; Jacobian calculation; Model selection; Uncertainty quantification

1. Introduction

The origin of fiducial inference can be traced back to R. A. Fisher (1922, 1925, 1930, 1933, 1935) who introduced the concept of a fiducial distribution for a parameter, and proposed the use of this fiducial distribution, in place of the Bayesian posterior distribution, for interval estimation of the parameter. In simple situations, especially in one parameter families of distributions, Fisher's fiducial intervals turned out to coincide with classical confidence intervals. For multi-parameter families of distributions, the fiducial approach led to confidence sets whose frequentist coverage probabilities were close to the claimed confidence levels but they were not exact in the repeated sampling frequentist sense. Fisher's proposal led to major discussions among the prominent statisticians of the mid-20th century (e.g., Jeffreys 1940; Stevens 1950; Tukey 1957; Lindley 1958; Fraser 1961a,b, 1966, 1968; Dempster 1966, 1968). Many of these discussions focused on the nonexactness of the confidence sets and also nonuniqueness of fiducial distributions. The latter part of the 20th century has seen only a handful of publications (Dawid, Stone, and Zidek 1973; Wilkinson 1977; Dawid and Stone 1982; Barnard 1995; Salome 1998) as the fiducial approach fell into disfavor and became a topic of historical interest only.

Since the mid-2000s, there has been a revival of interest in modern modifications of fiducial inference. This increase of interest demonstrated itself in both the number of different approaches to the problem and the number of researchers working on these problems, and is leading to an increasing number of publications in premier journals. The common thread for these approaches is a definition of inferentially meaningful probability statements about subsets of the parameter space without the need for subjective prior information.

These modern approaches include Dempster–Shafer theory (Dempster 2008; Edlefsen, Liu, and Dempster 2009) and recent (since 2010) related approach called *inferential models* (Martin, Zhang, and Liu 2010; Zhang and Liu 2011; Martin and Liu 2013, 2015a,c), which aims at provably conservative and efficient inference. While their philosophical approach to inference is different from ours, the resulting solutions are often mathematically closely related to the fiducial solutions presented here. Interested readers can learn about the inferential models approach to inference from the book by Martin and Liu (2015b). A somewhat different approach termed *confidence distributions* looks at the problem of obtaining an inferentially meaningful distribution on the parameter space from a purely frequentist point of view (Xie and Singh 2013). One of the main contributions of this approach is *fusion learning*: its ability to combine information from disparate sources with deep implications for meta analysis (Schweder and Hjort 2002; Singh, Xie, and Strawderman 2005; Xie, Singh, and Strawderman 2011; Hannig and

CONTACT Jan Hannig jan.hannig@unc.edu Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC 27599.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com//r/JASA.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

Xie 2012; Xie et al. 2013). Another related approach is based on higher order likelihood expansions and implied data-dependent priors (Fraser 2004, 2011; Fraser, Reid, and Wong 2005; Fraser and Naderi 2008; Fraser, Fraser, and Staicu 2009; Fraser et al. 2010). *Objective Bayesian inference*, which aims at finding nonsubjective model-based priors, is also part of this effort. Examples of recent breakthroughs related to *reference prior* and model selection are Bayarri et al. (2012), Berger (1992), Berger and Sun (2008), Berger, Bernardo, and Sun (2009), and Berger, Bernardo, and Sun (2012). Objective Bayesian inference is a very well-developed field but there is room for fiducial inference for many reasons. It often provides a good alternative both in terms of performance and ease of use, and the generalized fiducial distribution is never improper. Moreover, generalized fiducial inference and its various cousins are rapidly evolving and have the potential for uncovering deep and fundamental insights behind statistical inference. Finally, there are several other recent fiducial related works including Wang (2000), Xu and Li (2006), Veronese and Melilli (2015), and Taraldsen and Lindqvist (2013) who show how fiducial distributions naturally arise within a decision theoretical framework.

Arguably, *generalized fiducial inference* (GFI) has been on the forefront of the modern fiducial revival. It is motivated by the work of Tsui and Weerahandi (1989, 1991) and Weerahandi (1993, 1994, 1995) on *generalized confidence intervals* and the work of Chiang (2001) on the *surrogate variable method* for obtaining confidence intervals for variance components. The main spark came from the realization that there was a connection between these new procedures and fiducial inference. This realization evolved through a series of works (Iyer, Wang, and Mathew 2004; Patterson, Hannig, and Iyer 2004; Hannig, Iyer, and Patterson 2006b; Hannig 2009).

GFI defines a data-dependent measure on the parameter space by carefully using an inverse of a deterministic *data-generating equation* without the use of Bayes' theorem. The resulting *generalized fiducial distribution* (GFD) is a data-dependent distribution on the parameter space. GFD can be viewed as a distribution estimator (as opposed to a point or interval estimator) of the unknown parameter of interest. The resulting GFD when used to define approximate confidence sets is often shown in simulations to have very desirable properties, for example, conservative coverages but shorter expected lengths than competing procedures (E, Hannig, and Iyer 2008).

The strengths and limitations of the generalized fiducial approach are becoming better understood, see, especially, Hannig (2009, 2013). In particular, the asymptotic exactness of fiducial confidence sets, under fairly general conditions, was established in Hannig (2013), Hannig, Iyer, and Patterson (2006b), and Sonderegger and Hannig (2014). Higher order asymptotics of GFI was studied in Majumder and Hannig (2015). GFI has also been extended to prediction problems in Wang, Hannig, and Iyer (2012a). Model selection was introduced into the GFI paradigm in Hannig and Lee (2009). This idea was then further explored in classical setting in Wandler and Hannig (2011) and in the ultra high-dimensional regression in Lai, Hannig, and Lee (2015).

GFI has been proven useful in many practical applications. Earlier examples include bioequivalence (McNally, Iyer, and Mathew 2003; Hannig et al. 2006a), problems of metrology (Hannig, Wang, and Iyer 2003; Wang and Iyer 2005, 2006a,b; Hannig, Iyer, and Wang 2007; Wang, Hannig, and Iyer 2012b), and interlaboratory experiments and international key comparison experiments (Iyer, Wang, and Mathew 2004). It has also been applied to derive confidence procedures in many important statistical problems, such as variance components (E, Hannig, and Iyer 2008; Cisewski and Hannig 2012), maximum mean of a multivariate normal distribution (Wandler and Hannig 2011), multiple comparisons (Wandler and Hannig 2012a), extreme value estimation (Wandler and Hannig 2012b), mixture of normal and Cauchy distributions (Glagovskiy 2006), wavelet regression (Hannig and Lee 2009), and logistic regression and binary response models (Liu and Hannig 2016).

One main goal of this article is to deliver a concise introduction to GFI. Our intention is to provide a single location where the various developments of the last decade can be found. As a second goal of this article, some original work and refined results on GFI are also presented. Specifically, they are Definition 1 and Theorems 1, 3, and 4.

The rest of this article is organized as follows. Starting from Fisher's fiducial argument, Section 2 provides a complete description of GFI, including some new results. The issue of model selection within the GFI framework is discussed in Section 3. Section 4 concerns the use of GFI for discrete and discretized data, and Section 5 offers some practical advice on how to handle common computational challenges when applying GFI. Lastly, Section 5.1 provides some concluding remarks while technical details are relegated to the online appendix. The following website *http://anson.ucdavis.edu/∼tcmlee/GFiducial.html* contains computer code for many of the methods in this review.

2. The Switching Principle: Fisher's "Fiducial Argument" Extended

The idea underlying GFI is motivated by our understanding of Fisher's fiducial argument. GFI begins with expressing the relationship between the data, $\boldsymbol{Y}$, and the parameters, $\boldsymbol{\theta}$, as

$$\boldsymbol{Y} = \boldsymbol{G}(\boldsymbol{U}, \boldsymbol{\theta}), \tag{1}$$

where $\boldsymbol{G}(\cdot, \cdot)$ is a deterministic function termed the *data-generating equation*, and $\boldsymbol{U}$ is the random component of this data-generating equation whose distribution is independent of parameters and completely known.

The data $\boldsymbol{Y}$ are assumed to be created by generating a random variable $\boldsymbol{U}$ and plugging it into the data-generating Equation (1). For example, a single observation from $N(\mu, 1)$ distribution can be written as $Y = \mu + U$, where $\boldsymbol{\theta} = \mu$ and U is $N(0, 1)$ random variable.

For simplicity, this subsection only considers the simple case where the data-generating Equation (1) can be inverted and the inverse $Q_{\boldsymbol{y}}(\boldsymbol{u}) = \boldsymbol{\theta}$ exists for any observed $\boldsymbol{y}$ and for any arbitrary $\boldsymbol{u}$. Fisher's *fiducial argument* leads one to define the fiducial distribution for $\boldsymbol{\theta}$ as the distribution of $Q_{\boldsymbol{y}}(\boldsymbol{U}^\star)$ where $\boldsymbol{U}^\star$ is an independent copy of $\boldsymbol{U}$. Equivalently, a sample from the fiducial distribution of $\boldsymbol{\theta}$ can be obtained by generating $\boldsymbol{U}_i^\star$, $i = 1, \ldots, N$ and using $\boldsymbol{\theta}_i^\star = Q_{\boldsymbol{y}}(\boldsymbol{U}_i^\star)$. Estimates and confidence intervals for

$\boldsymbol{\theta}$ can be obtained based on this sample. In the $N(\mu, 1)$ example, $Q_y(u) = y - u$ and the fiducial distribution is therefore the distribution of $y - U^\star \sim N(y, 1)$.

Example 1. Consider the mean and sample variance $\boldsymbol{Y} = (\bar{Y}, S^2)$ computed from n independent $N(\mu, \sigma^2)$ random variables, where μ and σ^2 are parameters to be estimated. A natural data-generating equation for $\boldsymbol{Y}$ is

$$\bar{Y} = \mu + \sigma U_1 \text{ and } S^2 = \sigma^2 U_2,$$

where U_1, U_2 are independent with $U_1 \sim N(0, n^{-1})$ and $U_2 \sim \text{Gamma}((n-1)/2, (n-1)/2)$.

The inverse $Q_{\boldsymbol{y}}(\boldsymbol{u}) = (\bar{y} - s\,u_1/\sqrt{u_2}, s^2/u_2)$. Consequently, for any observed value $\bar{y}$ and s, and an independent copy of $\boldsymbol{U}$, denoted as $\boldsymbol{U}^\star$, the distribution of $\mu^\star = \bar{y} - s\,U_1^\star/\sqrt{U_2^\star}$ is the marginal fiducial distribution of μ. The equal tailed set of 95% fiducial probability is $(\bar{y} - ts/\sqrt{n}, \bar{y} + ts/\sqrt{n})$ where t is the 0.025 critical value of the t-distribution with $n-1$ degrees of freedom, which is the classical 95% confidence interval for μ.

Remark 1. Generalized fiducial distribution is a data-dependent measure on the parameter space. It is mathematically very similar to Bayesian posteriors and can be used in practice in a similar fashion. For example, the median (or mean) of the GFD can be used as a point estimator. More importantly, certain sets of fiducial probability $1-\alpha$ can be used as approximate $(1-\alpha)100\%$ confidence sets, see Theorem 3. A nested collection of such approximate confidence sets at all confidence levels can also be inverted for a use as an approximate p-value (Hannig 2009; Xie and Singh 2013).

A useful graphical tool for visualizing GFDs is the confidence curve of Birnbaum (1961). If $R(\theta|\mathbf{x})$ is the distribution (or survival) function of a marginal fiducial distribution, the confidence curve is defined as $\text{CV}(\theta) = 2|R(\theta|\mathbf{x}) - 0.5|$. On a plot of $\text{CV}(\theta)$ versus θ, a line across the height (y-axis) of α, for any $0 < \alpha < 1$, intersects with the confidence curve at two points, and these two points correspond (on x-axis) to an α level, equal tailed, two-sided confidence interval for θ. Thus, a confidence curve is a graphical device that shows confidence intervals of all levels. The minimum of a confidence curve is the median of the fiducial distribution which is the recommended point estimator. Figure 1 shows an example of a confidence curve for a dataset generated from $U(\theta, \theta^2)$ distribution of Example 4.

Remark 2. We have made a conscious choice to eschew philosophical controversies throughout the development of GFI. However, we find it inevitable to make at least some philosophical comments at this point:

1. The idea behind GFD is very similar to the idea behind the likelihood function: what is the chance of observing my data if any given parameter was true. The added value of GFD is that it provides likelihood function with an appropriate Jacobian obtaining a proper probability distribution on the parameter space, see (4) below.
2. GFD does not presume that the parameter is random. Instead it should be viewed as a distribution estimator (rather than a point or interval estimator) of the fixed true parameter. To validate this distribution estimator in a specific example, we then typically demonstrate good small sample performance by simulation and prove good large sample properties by asymptotic theorems.
3. From a Bayesian point of view, Bayes' theorem updates the distribution of $\boldsymbol{U}$ after the data are observed. However, when no prior information is present, changing the distribution of $\boldsymbol{U}$ only by restricting it to the set "there is at least one $\boldsymbol{\theta}$ solving the equation $\boldsymbol{y} = \boldsymbol{G}(\boldsymbol{U}, \boldsymbol{\theta})$" seems to us as a reasonable choice (see the next section). Arguably, this so-called "continuing to regard" assumption has been behind most of the philosophical controversies surrounding fiducial inference in the past.

2.1. A Refined Definition of Generalized Fiducial Distribution

The inverse to Equation (1) does not exist for two possible reasons. Either, there is more than one $\boldsymbol{\theta}$ for some value of $\boldsymbol{y}$ and $\boldsymbol{u}$, or there is no $\boldsymbol{\theta}$ satisfying $\boldsymbol{y} = \boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta})$. The first situation can be dealt with by using the mechanics of Dempster–Shafer calculus (Dempster 2008). A more practical solution is to select one of the several solutions using a possibly random mechanism. In Section 4, we will review theoretical results that showed that the uncertainty due to multiple solutions has, in many parametric problems, only a second-order effect on statistical inference.

For the second situation, Hannig (2009) suggested removing the values of $\boldsymbol{u}$ for which there is no solution from the sample space and then renormalizing the probabilities, that is, using the distribution of $\boldsymbol{U}$ conditional on the event $\mathcal{U}_{\boldsymbol{y}} = \{\boldsymbol{u} : \boldsymbol{y} = \boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta}), \text{ for some } \boldsymbol{\theta}\}$. The rationale for this choice is that we know that the observed data $\boldsymbol{y}$ were generated using some fixed unknown $\boldsymbol{\theta}_0$ and $\boldsymbol{u}_0$, that is, $\boldsymbol{y} = \boldsymbol{G}(\boldsymbol{\theta}_0, \boldsymbol{u}_0)$. The values of $\boldsymbol{u}$ for which $\boldsymbol{y} = \boldsymbol{G}(\cdot, \boldsymbol{u})$ does not have a solution could not be the true $\boldsymbol{u}_0$ hence only the values of $\boldsymbol{u}$ for which there is a solution should be considered in the definition of the *generalized fiducial distribution* (an exception to this suggestion is in cases where the parameter space is in some way constrained. In this case, it is often beneficial to extend the parameter space, perform the inversion in the extended space, and then project to the boundary of the constrained parameter space. A good example of such a situation is the variance component model where variances are constrained to be greater than or equal to zero (E, Hannig, and Iyer 2009; Cisewski and Hannig 2012)). However, $\mathcal{U}_{\boldsymbol{y}}$, the set of $\boldsymbol{u}$ for which the solution exists, has probability zero for most problems involving absolutely continuous random variables. Conditioning on such a set of probability zero will therefore lead to nonuniqueness due to the Borel paradox (Casella and Berger 2002, sec. 4.9.3).

Hannig (2013) proposed an attractive interpretation of the conditional distribution by limit of discretizations. Here, we generalize this approach slightly. Throughout this manuscript, $\boldsymbol{U}^\star$ denotes an independent copy of $\boldsymbol{U}$ and $\boldsymbol{\theta}^\star$ denotes a random variable taking values in the parameter space $\boldsymbol{\Theta}$.

To define GFD, we need to interpret the ill-defined conditional distribution of $\boldsymbol{U}^\star \mid \boldsymbol{U}^\star \in \mathcal{U}_{\boldsymbol{y}}$. To do that we "fatten up" the manifold $\mathcal{U}_{\boldsymbol{y}}$ by ϵ so that the enlarged set $\mathcal{U}_{\boldsymbol{y},\epsilon} = \{\boldsymbol{u} : \|\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta})\| \le \epsilon, \text{ for some } \boldsymbol{\theta}\}$ has positive probability and the conditional distribution of $\boldsymbol{U}^\star \mid \boldsymbol{U}^\star \in \mathcal{U}_{\boldsymbol{y},\epsilon}$ is well defined. Finally, the fattening needs to be done in a consistent way so that the

limit of conditional distributions as $\epsilon \to 0$ is well defined. This leads to the following definition:

Definition 1. A probability measure on the parameter space $\boldsymbol{\Theta}$ is called a generalized fiducial distribution (GFD) if it can be obtained as a weak limit:

$$\lim_{\epsilon \to 0} \left[\arg\min_{\boldsymbol{\theta}^\star} \|\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{U}^\star, \boldsymbol{\theta}^\star)\| \;\middle|\; \min_{\boldsymbol{\theta}^\star} \|\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{U}^\star, \boldsymbol{\theta}^\star)\| \le \epsilon \right]. \tag{2}$$

If there are multiple minimizers $\arg\min_{\boldsymbol{\theta}^\star} \|\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{U}^\star, \boldsymbol{\theta}^\star)\|$, one selects one of them (potentially at random). Notice that the conditioning in (2) is modifying the distribution of $\boldsymbol{U}^\star$ to only consider values for which an approximate inverse to $\boldsymbol{G}$ exists.

Remark 3. Definition 1 illuminates the relationship between GFD and Approximate Bayesian Computations (ABC; Beaumont, Zhang, and Balding 2002). In an idealized ABC, one generates first an observation $\boldsymbol{\theta}^*$ from the prior, then generates a new sample using a data-generating equation $\boldsymbol{y}^\star = \boldsymbol{G}(\boldsymbol{U}^\star, \boldsymbol{\theta}^\star)$ and compares the generated data with the observed data $\boldsymbol{y}$. If the observed and generated datasets are close (e.g., $\|\boldsymbol{y} - \boldsymbol{y}^\star\| \le \epsilon$), the generated $\boldsymbol{\theta}^\star$ is accepted, otherwise it is rejected and the procedure is repeated. If the measure of closeness is a norm, it is easy to see that when $\epsilon \to 0$ the weak limit of the ABC distribution is the posterior distribution.

On the other hand, when defining GFD one generates $\boldsymbol{U}^\star$, finds a best-fitting $\boldsymbol{\theta}^\star = \arg\min_{\boldsymbol{\theta}^\star} \|\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{U}^\star, \boldsymbol{\theta}^\star)\|$, computes $\boldsymbol{y}^\star = \boldsymbol{G}(\boldsymbol{U}^\star, \boldsymbol{\theta}^\star)$, again accepts $\boldsymbol{\theta}^\star$ if $\|\boldsymbol{y} - \boldsymbol{y}^\star\| \le \epsilon$, and rejects otherwise.

In either approach, an artificial dataset $\boldsymbol{y}^\star = \boldsymbol{G}(\boldsymbol{U}^\star, \boldsymbol{\theta}^\star)$ is generated and compared to the observed data. The main difference is that the Bayes posterior simulates the parameter $\boldsymbol{\theta}^\star$ from the prior while GFD uses the best-fitting parameter.

Remark 4. The GFD defined in (2) is not unique as it depends on both the data-generating Equation (1), the norm used in (2) and the minimizer $\boldsymbol{\theta}^\star$ chosen. Let $\boldsymbol{U}^\star$ be an independent copy of $\boldsymbol{U}$ and let for any measurable set A, $V[A]$ be a rule selecting a possibly random element of the closure of the set $\bar{A}$. When the probability $P(\exists \boldsymbol{\theta}^\star,\ \boldsymbol{y} = \boldsymbol{G}(\boldsymbol{U}^\star, \boldsymbol{\theta}^\star)) > 0$ then the limit (2) is the conditional distribution

$$V[\{\boldsymbol{\theta}^\star : \boldsymbol{y} = \boldsymbol{G}(\boldsymbol{U}^\star, \boldsymbol{\theta}^\star)\}] \mid \{\exists \boldsymbol{\theta}^\star,\ \boldsymbol{y} = \boldsymbol{G}(\boldsymbol{U}^\star, \boldsymbol{\theta}^\star)\}.$$

This is an older definition of GFD that can be found in Hannig (2009, 2013).

The next subsection offers a useful computational formula for evaluating (2).

2.2. A User Friendly Formula for Generalized Fiducial Distribution

While Definition (2) for GFD is conceptually appealing and very general, it is not immediately clear how to compute the limit in many practical situations. In a less general setup using the l^∞ norm, Hannig (2013) derived a closed form of the limit in (2) applicable to many practical situations. Here, we provide a generalization of this result, which is applicable in most situations where the data follow a continuous distribution.

Assume that the parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$ is p-dimensional, the data $\boldsymbol{x} \in \mathbb{R}^n$ are n dimensional. The following theorem provides a useful computational formula.

Theorem 1. Suppose Assumptions A.1 to A.3 stated in Appendix A. Then the limiting distribution in (2) has a density

$$r(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}, \boldsymbol{\theta}) J(\boldsymbol{y}, \boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} f(\boldsymbol{y}, \boldsymbol{\theta}') J(\boldsymbol{y}, \boldsymbol{\theta}')\, d\boldsymbol{\theta}'}, \tag{3}$$

where $f(\boldsymbol{y}, \boldsymbol{\theta})$ is the likelihood and the function

$$J(\boldsymbol{y}, \boldsymbol{\theta}) = D\left(\frac{d}{d\boldsymbol{\theta}} \boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta}) \Big|_{\boldsymbol{u} = \boldsymbol{G}^{-1}(\boldsymbol{y}, \boldsymbol{\theta})} \right). \tag{4}$$

If (i) $n = p$ then $D(A) = |\det A|$. Otherwise the function $D(A)$ depends on the norm used; (ii) the l_∞ norm gives $D(A) = \sum_{\mathbf{i}=(i_1,\ldots,i_p)} |\det(A)_{\mathbf{i}}|$; (iii) under an additional Assumption A.4 the l_2 norm gives $D(A) = (\det A^\top A)^{1/2}$.

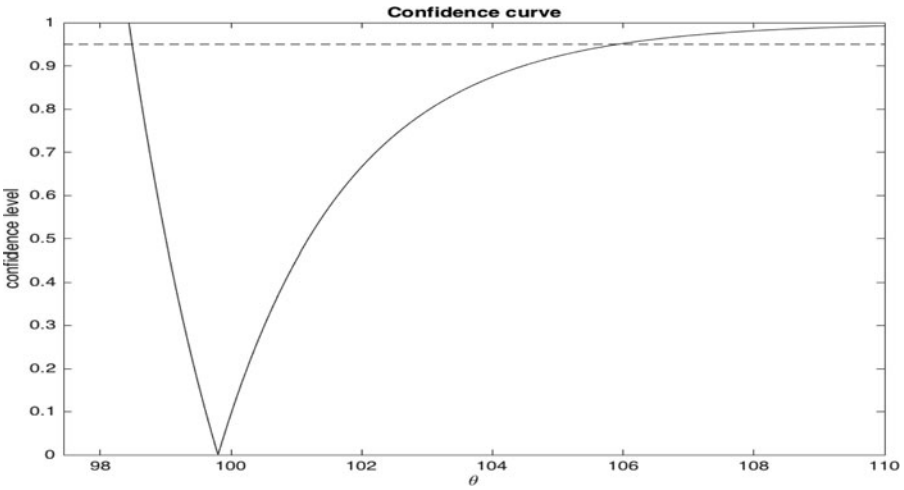

Figure 1. Confidence curve of a GFD for parameter θ based on sample of size 25 from $U(\theta, \theta^2)$ with $\theta = 100$. The minimum and maximum values of the sample used to generate the GFD are 281.1 and 9689.7, respectively. The interval between the two points where the dotted line intersects the confidence curve (98.49, 105.92) is the approximate 95% confidence interval. The minimum of the confidence curve is the median of the generalized fiducial distribution. Its value of 99.8 provides a natural point estimator.

In (ii) the sum spans over $\binom{n}{p}$ of p-tuples of indexes $\boldsymbol{i} = (1 \le i_1 < \cdots < i_p \le n)$. For any $n \times p$ matrix A, the submatrix $(A)_{\boldsymbol{i}}$ is the $p \times p$ matrix containing the rows $\boldsymbol{i} = (i_1, \ldots, i_p)$ of A.

There is a slight abuse of notation in (3) as $r(\boldsymbol{\theta}|\boldsymbol{y})$ is not a conditional density in the usual sense. Instead, we are using this notation to remind the reader that the fiducial density depends on the fixed observed data.

Cases (i) and (ii) are a simple consequence of results in Hannig (2013). The formula in (iii) was independently proposed in Fraser et al. (2010) based on arguments related to tangent exponential families without being recognized as a fiducial distribution. The proof is in Appendix A. The ease of use of (4) will be demonstrated on several examples in the next subsection. The rest of this subsection discusses the effects of various transformations.

Remark 5. Just like posterior computed using Jeffreys' prior, GFD is invariant under smooth reparameterizations.

This assertion has been shown for smooth transformation by chain rule in Hannig (2013). However, this property is general and follows directly from (2), since for an appropriate selection of minimizers and any one-to-one function $\boldsymbol{\theta} = \phi(\boldsymbol{\eta})$

$$\phi\left(\arg\min_{\boldsymbol{\eta}^\star} \|\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{U}^\star, \phi(\boldsymbol{\eta}^\star))\|\right) = \arg\min_{\boldsymbol{\theta}^\star} \|\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{U}^\star, \boldsymbol{\theta}^\star)\|.$$

Remark 6. GFD could change with transformations of the data-generating equation.

Assume that the observed dataset has been transformed with a one-to-one smooth transformation $\boldsymbol{Z} = \boldsymbol{T}(\boldsymbol{Y})$. Using the chain rule, we see that the GFD based on this new data-generating equation and with observed data $\boldsymbol{z} = \boldsymbol{T}(\boldsymbol{y})$ is the density (3) with the Jacobian function (4) simplified to

$$J_T(\boldsymbol{z}, \boldsymbol{\theta}) = D\left(\frac{d}{d\boldsymbol{y}}\boldsymbol{T}(\boldsymbol{y}) \cdot \frac{d}{d\boldsymbol{\theta}}\boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta})\Big|_{\boldsymbol{u}=\boldsymbol{G}^{-1}(\boldsymbol{y},\boldsymbol{\theta})}\right). \tag{5}$$

Notice that for simplicity we write $\boldsymbol{y}$ instead of $T^{-1}(\boldsymbol{z})$ in (5).

For completeness, we recall the well-known fact that the likelihood based on $\boldsymbol{z} = \boldsymbol{T}(\boldsymbol{y})$ satisfies

$$f_T(\boldsymbol{z}|\boldsymbol{\theta}) = f(\boldsymbol{y}|\boldsymbol{\theta})\left|\det\left(\frac{d\boldsymbol{T}}{d\boldsymbol{y}}\right)\right|^{-1}. \tag{6}$$

The second term in the right-hand side of (6) is a constant and does not affect the GFD with the exception of model selection considerations in Section 3.

As can be seen from the above calculation, GFD will usually change with transformation of the data. An important exception is when the number of observations and number of parameters are equal, that is, $n = p$. Indeed, by careful evaluation of (4), (5), and (6), we see that for $\boldsymbol{z} = \boldsymbol{T}(\boldsymbol{y})$ we have $J(\boldsymbol{y}, \boldsymbol{\theta}) f_Y(\boldsymbol{y}|\boldsymbol{\theta}) = J_T(\boldsymbol{z}, \boldsymbol{\theta}) f_T(\boldsymbol{z}|\boldsymbol{\theta})$ and the GFD is unchanged.

Example 2. Consider the following important transformation. Let $\boldsymbol{Z} = (\boldsymbol{S}, \boldsymbol{A})^\top$ be one-to-one smooth transformation, where $\boldsymbol{S}$ is a p-dimensional statistic and $\boldsymbol{A}$ is an ancillary statistic. Let $\boldsymbol{s} = \boldsymbol{S}(\boldsymbol{y})$ and $\boldsymbol{a} = \boldsymbol{A}(\boldsymbol{y})$ be the observed values. Since $d\boldsymbol{A}/d\boldsymbol{\Theta} = 0$, the function D in (5) is the absolute value of the determinant of the $p \times p$ nonzero submatrix:

$$J(\boldsymbol{z}, \boldsymbol{\theta}) = \left|\det\left(\frac{d}{d\boldsymbol{\theta}}\boldsymbol{S}(\boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta}))\Big|_{\boldsymbol{u}=\boldsymbol{G}^{-1}(\boldsymbol{y},\boldsymbol{\theta})}\right)\right|. \tag{7}$$

Next, denote the solution of the equation $\boldsymbol{s} = \boldsymbol{S}(\boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta}))$ by $Q_{\boldsymbol{s}}(\boldsymbol{u}) = \boldsymbol{\theta}$. A straightforward calculation shows that the fiducial density (3) with (7) is the conditional distribution of $Q_{\boldsymbol{s}}(\boldsymbol{U}^\star) \mid \boldsymbol{A}(\boldsymbol{U}^\star) = \boldsymbol{a}$, the GFD based on $\boldsymbol{S}$ conditional on the observed ancillary $\boldsymbol{A} = \boldsymbol{a}$, see Birnbaum (1962) and Iyer and Patterson (2002).

2.3. Two Examples

In this section, we will consider two examples, linear regression and uniform distribution. In the first case, the GFD is the same as Bayes posterior with respect to the independence Jeffreys' prior while in the second the GFD is not a Bayes posterior with respect to any prior (that is not data dependent).

Example 3 (Linear Regression). Express linear regression using the data-generating equation

$$\boldsymbol{Y} = G(\boldsymbol{U}, \boldsymbol{\theta}) = \boldsymbol{X}\beta + \sigma\boldsymbol{U},$$

where $\boldsymbol{Y}$ is the dependent variables, $\boldsymbol{X}$ is the design matrix, $\boldsymbol{\theta} = (\beta, \sigma)$ are the unknown parameters, and $\boldsymbol{U}$ is a random vector with known density $f(\boldsymbol{u})$ independent of any parameters.

To compute GFD, simply notice that $\frac{d}{d\boldsymbol{\theta}}\boldsymbol{G}(\boldsymbol{U}, \boldsymbol{\theta}) = (\boldsymbol{X}, \boldsymbol{U})$, $\boldsymbol{U} = \sigma^{-1}(\boldsymbol{y} - \boldsymbol{X}\beta)$. From here the Jacobian in (4) using the l_∞ norm simplifies to

$$J_\infty(\boldsymbol{y}, \boldsymbol{\theta}) = \sigma^{-1} \sum_{\substack{\boldsymbol{i}=(i_1,\ldots,i_p) \\ 1 \le i_1 < \cdots < i_p \le n}} |\det(\boldsymbol{X}, \boldsymbol{Y})_{\boldsymbol{i}}|$$

and the density of GFD is

$$r(\beta, \sigma|\boldsymbol{y}) \propto \sigma^{-n-1} f(\sigma^{-1}(\boldsymbol{Y} - \boldsymbol{X}\beta)).$$

This coincides with the Bayesian solution using the independence Jeffreys' prior (Yang and Berger 1997).

The J function has a more compact form when using the l_2 norm. In particular by Cauchy–Binet formula, we see that $\det((\boldsymbol{X}, \boldsymbol{y} - \boldsymbol{X}\beta)^\top(\boldsymbol{X}, \boldsymbol{y} - \boldsymbol{X}\beta))$ is invariant in β. By selecting $\beta = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$, we immediately obtain

$$J_2(\boldsymbol{y}, \boldsymbol{\theta}) = \sigma^{-1}|\det(\boldsymbol{X}^\top\boldsymbol{X})|^{\frac{1}{2}}\mathrm{RSS}^{\frac{1}{2}},$$

where RSS is the residual sum of squares. As the two Jacobian functions differ only by a constant, the GFD is unchanged.

As a special case, the GFD for the location-scale model $\boldsymbol{X} = \boldsymbol{1}$, the l_∞ Jacobian is $J_\infty(\boldsymbol{y}, \boldsymbol{\theta}) = \sigma^{-1}\sum_{i<j}|Y_i - Y_j|$ while the l_2 Jacobian becomes $J_2(\boldsymbol{y}, \boldsymbol{\theta}) = \sigma^{-1}n\hat{\sigma}_n$, where $\hat{\sigma}_n$ is the maximum likelihood estimator of σ.

Example 4 (Uniform $U\{a(\theta) - b(\theta), a(\theta) + b(\theta)\}$). As a second example, we will study a very irregular model. The reference prior for this model is complicated and has been obtained as Theorem 8 in Berger, Bernardo, and Sun (2009).

Express the observed data using the following data-generating equation

$$Y_i = a(\theta) + b(\theta)U_i, \quad U_i \text{ iid } U(-1, 1).$$

Simple computations give $\frac{d}{d\theta}\boldsymbol{G}(\boldsymbol{u},\theta) = a'(\theta) + b'(\theta)\boldsymbol{U}$ with $\boldsymbol{U} = b^{-1}(\theta)(\boldsymbol{Y} - a(\theta))$. If $a'(\theta) > |b'(\theta)|$, (4) simplifies to

$$J_1(\boldsymbol{y},\theta) = \sum_{i=1}^{n} \left| a'(\theta) + \{\log b(\theta)\}'\{y_i - a(\theta)\} \right| = n[a'(\theta) - a(\theta)\{\log b(\theta)\}' + \bar{y}_n\{\log b(\theta)\}']. \quad (8)$$

We used $a'(\theta) > |b'(\theta)|$ only to show that the terms inside the absolute values below are all positive. However, we remark that under this assumption both $a(\theta) - b(\theta)$ and $a(\theta) + b(\theta)$ are strictly increasing, continuous functions of θ.

With the above, the GFD is then

$$r(\theta|\boldsymbol{y}) \propto \frac{a'(\theta) - a(\theta)\{\log b(\theta)\}' + \bar{y}_n\{\log b(\theta)\}'}{b(\theta)^n} \times I_{\{a(\theta)-b(\theta)<y_{(1)} \ \&\ a(\theta)+b(\theta)>y_{(n)}\}}. \quad (9)$$

As an alternative fiducial solution, consider a transformation to the minimal sufficient and ancillary inspired by of the Example 2. $\boldsymbol{Z} = \{h_1(Y_{(1)}), h_2(Y_{(n)}), (\boldsymbol{Y} - Y_{(1)})/(Y_{(n)} - Y_{(1)})\}^\top$. We selected the transformations h_i so that their inverse $h_1^{-1}(\theta) = EY_{(1)} = a(\theta) - b(\theta)(n-1)/(n+1)$ and $h_2^{-1}(\theta) = EY_{(n)} = a(\theta) + b(\theta)(n-1)/(n+1)$. There are only two nonzero terms in (5) and consequently

$$J_2(\boldsymbol{y},\theta) = (w_1 + w_2)\left[a'(\theta) - a(\theta)\{\log b(\theta)\}' + \frac{w_1 y_{(1)} + w_2 y_{(n)}}{w_1 + w_2}\{\log b(\theta)\}'\right], \quad (10)$$

where $w_1 = h_1'(y_{(1)})$ and $w_2 = h_2'(y_{(n)})$.

We performed a simulation study for the particular case of $U(\theta,\theta^2)$; $a(\theta) = \theta$, $b(\theta) = \theta^2 - \theta$. For this model, the likelihood is $f(\boldsymbol{y}|\theta) = \{\theta(\theta-1)\}^{-n} I_{(y_{(n)}^{1/2}, y_{(1)})}(\theta)$ and the Jacobians are

$$J_1(\boldsymbol{y},\theta) = n\frac{\bar{y}(2\theta-1) - \theta^2}{\theta(\theta-1)} \quad \text{and}$$

$$J_2(\boldsymbol{y},\theta) = \frac{(w_1 y_{(1)} + w_2 y_{(n)})(2\theta-1) - (w_1+w_2)\theta^2}{\theta(\theta-1)},$$

where

$$w_1 = \frac{1+n}{\sqrt{n^2 + 4(1+n)y_{(1)}}}, \quad w_2 = \frac{1+n}{\sqrt{1 + 4n(1+n)y_{(n)}}}.$$

An example of confidence curve for a GFD based on (10) is in Figure 1.

We compared the performance of the two fiducial distributions to the Bayesian posteriors with the reference prior $\pi(\theta) = \frac{(2\theta-1)}{\theta(\theta-1)} e^{\psi\left(\frac{2\theta}{2\theta-1}\right)}$ (Berger, Bernardo, and Sun 2009) ($\psi(x)$ is the digamma function defined by $\psi(z) = \frac{d}{dz}\log(\Gamma(z))$ for $z > 0$) and flat prior $\pi(\theta) = 1$.

For all the combinations of $n = 1, 2, 3, 4, 5, 10, 20, 100, 250$ and $\theta = 1.5, 2, 5, 10, 25, 50, 100, 250$ we analyzed 16,000 independent datasets. Based on this we found the empirical coverage of the 2.5%, 5%, 50%, 95%, 97.5% upper confidence bounds. The results are summarized in Figure 2. We observed that the simple GFD (denoted by F1 in the figures), the alternative GFD based on minimal sufficient statistics (F2) and the reference prior Bayes posterior (BR) maintain stated coverage for all parameter settings. However, the flat prior Bayes posterior (B1) does not have a satisfactory coverage, with the worst departures from stated coverage observed for small n and large θ.

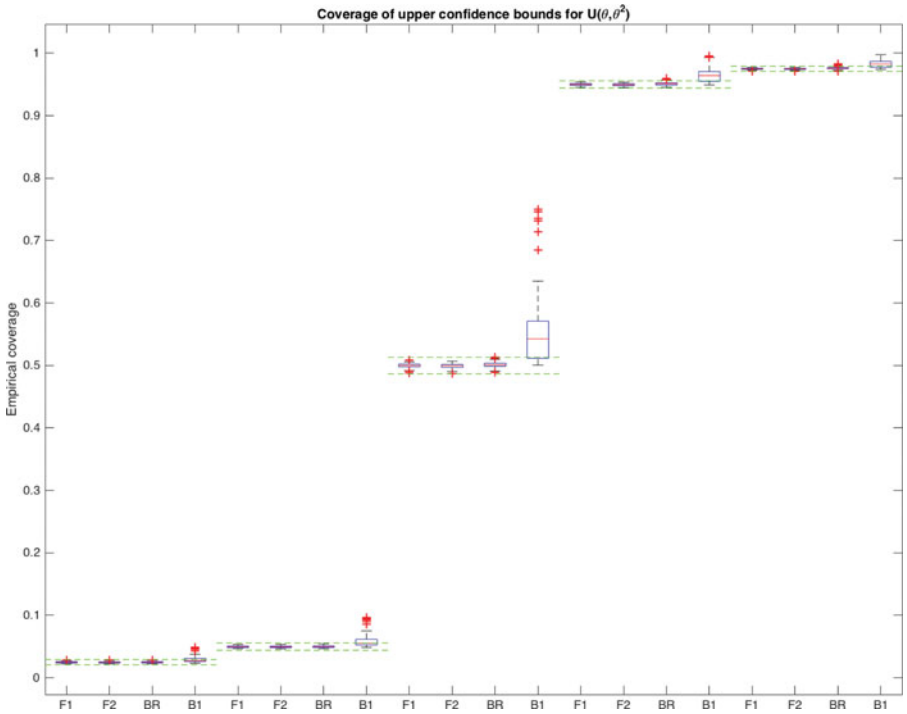

Figure 2. Boxplots of empirical coverages of the 2.5%, 5%, 50%, 95%, 97.5% upper confidence bounds for the simple GFD (F1), GFD based on minimal sufficient statistics (F2), reference Bayes (BR), and flat prior Bayes (B1) for all 72 parameter settings. The broken lines provide bounds on random fluctuations of the empirical coverages showing that F1, F2, and BR maintain stated coverage while B1 does not.

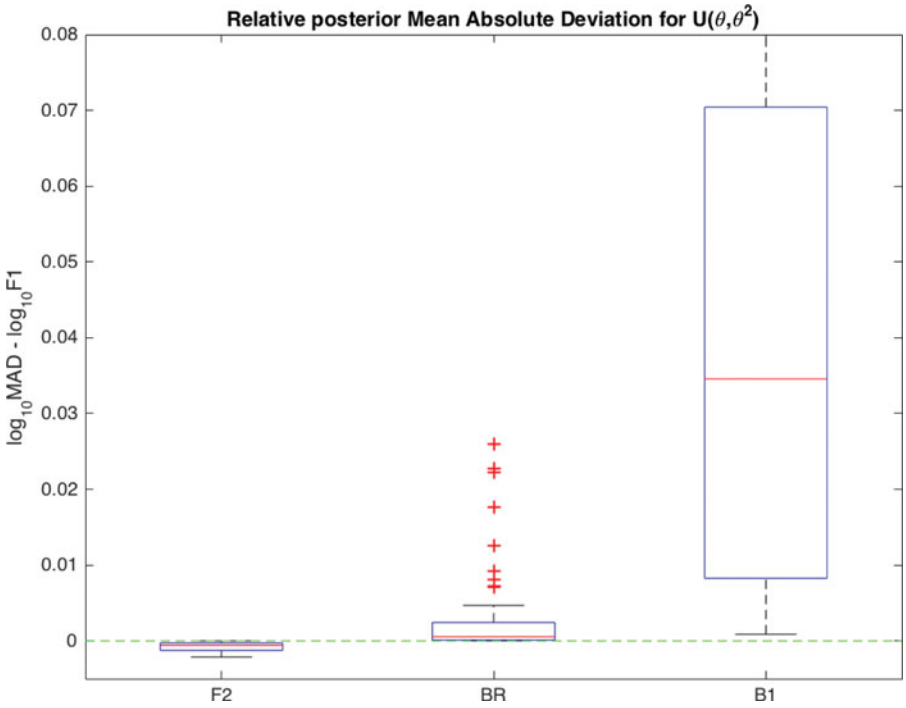

Figure 3. Boxplots of $\log_{10}(\text{MAD})$ of GFD based on minimal sufficient statistics (F2), reference Bayes (BR), and flat prior Bayes minus the $\log_{10}(\text{MAD})$ of the simple fiducial (F1) averaged over 16,000 simulated datasets. Each 72 parameter combinations provide one data point for the boxplots. Positive values mean that GFD F1 concentrates closer to the truth; consequently, F2 is the best in this metric.

For each dataset, we have also measured the mean absolute deviation from the true parameter (MAD) of each of the GFD and posteriors, that is, $\text{MAD} = \int |\theta - \theta_0| r(\theta|\boldsymbol{y}) d\theta$. To aid in comparison, we compute the difference of the $\log_{10}(\text{MAD})$ of F2, BR, and B1 minus the $\log_{10}(\text{MAD})$ of F1 on each dataset. A positive (resp., negative) value of the difference signifies that the posterior is concentrated further from (resp., closer to) the truth than the simple fiducial. These relative MADs are then averaged across the 16,000 simulated datasets for all the parameter settings and reported in Figure 3. We observe that F2 is better than F1, which is better than BR though the absolute value of the difference is relatively small. B1 is not competitive.

2.4. Theoretical Results

This section discusses asymptotic properties for GFI. We hope that the material included here will be useful for the study of GFD in future practical problems.

First, we present a Bernstein–von Mises theorem for GFD, which provides theoretical guarantees of asymptotic normality and asymptotic efficiency. It also guarantees in conjunction with Theorem 3 that appropriate sets of fiducial probability $1-\alpha$ are indeed approximate $1-\alpha$ confidence sets. An early version of this theorem can be found in Hannig (2009). Here, we will state a more general result due to Sonderegger and Hannig (2014).

Assume that we are given a random sample of independent observations $Y_1, \ldots, Y_n$ with data-generating equation $Y_i = \boldsymbol{G}(\boldsymbol{\theta}, U_i)$, with U_i iid $U(0,1)$. This data-generating equation leads to the Jacobian function (4) that is a U-statistic. This realization makes the GFD amenable to theoretical study.

Asymptotic normality of statistical estimators usually relies on a set of technical assumptions and GFD is no exception. To succinctly state the theorem, we denote the rescaled density of GFD by $r^*(\boldsymbol{s}|\boldsymbol{y}) = n^{-1/2} r(n^{-1/2}\boldsymbol{s} + \hat{\boldsymbol{\theta}}|\boldsymbol{y})$, where $\hat{\boldsymbol{\theta}}$ is the consistent maximum likelihood estimator (MLE).

Theorem 2 (Sonderegger and Hannig 2014). Under Assumptions B.1 to B.4 in Appendix B

$$\int_{\mathbb{R}^p} \left| r^*(\boldsymbol{s}|\boldsymbol{y}) - \frac{\sqrt{\det|I(\boldsymbol{\theta}_0)|}}{\sqrt{2\pi}} e^{-\boldsymbol{s}^T I(\boldsymbol{\theta}_0)\boldsymbol{s}/2} \right| d\boldsymbol{s} \xrightarrow{P_{\boldsymbol{\theta}_0}} 0.$$

One application of GFD is to take sets of $1-\alpha$ fiducial probability and use them as approximate confidence intervals. Next we state conditions under which this is valid.

Assumption 1. Let us consider a sequence of datasets $\boldsymbol{Y}_n$ generated using fixed parameters $\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n$ with corresponding data-dependent measures (such data-dependent measures can be, for example, GFD s, Bayes posteriors, or confidence distributions) on the parameter space $R_{n,\boldsymbol{Y}_n}$. We will assume that these converge to a limiting fiducial model in the following way:

1. There is a sequence of measurable functions t_n of $\boldsymbol{Y}_n$ so that $t_n(\boldsymbol{Y}_n)$ converges in distribution to some random variable $\boldsymbol{T}$.
2.
   (a) The $\boldsymbol{T}$ from Part 1 can be decomposed into $\boldsymbol{T} = (\boldsymbol{T}_1, \boldsymbol{T}_2)$ and there is a limiting data-generating equation $\boldsymbol{T}_1 = \boldsymbol{H}_1(\boldsymbol{V}_1, \boldsymbol{\xi})$, $\boldsymbol{T}_2 = \boldsymbol{H}_2(\boldsymbol{V}_2)$, where $\boldsymbol{V} = (\boldsymbol{V}_1, \boldsymbol{V}_2)$ has a fully known distribution independent of the parameter $\boldsymbol{\xi} \in \boldsymbol{\Xi}$. The distribution of $\boldsymbol{T}$ is obtained from the limiting data-generating equation using $\boldsymbol{\xi}_0$.
   (b) The equation $\boldsymbol{H}_1$ is one-to-one if viewed as a function (possibly implicit) for any combination of $\boldsymbol{\xi}$, $\boldsymbol{v}_1$, $\boldsymbol{t}_1$, where one is held fixed, one taken as a dependent, and one taken as an independent variable. The

equation $\boldsymbol{H}_2$ is one to one. Consequently, the limiting GFD defined by (2) is the conditional distribution $Q_{t_1}(\boldsymbol{V}_1^\star) \mid \boldsymbol{H}_2(\boldsymbol{V}_2^\star) = \boldsymbol{t}_2$, where $Q_{t_1}(\boldsymbol{v}_1) = \boldsymbol{\xi}$ is the solution of $\boldsymbol{t}_1 = \boldsymbol{H}_1(\boldsymbol{v}_1, \boldsymbol{\xi})$. Denote this conditional measure by $R_{\boldsymbol{t}}$.

(c) For any open set $C \subset \boldsymbol{\Xi}$ and limiting data $\boldsymbol{t}$, the limiting fiducial probability of the boundary $R_{\boldsymbol{t}}(\partial C) = 0$.

3. There are homeomorphic injective mappings Ξ_n from $\boldsymbol{\Theta}_n$ into $\boldsymbol{\Xi}$ so that
   (a) $\Xi_n(\boldsymbol{\theta}_{n,0}) = \boldsymbol{\xi}_0$;
   (b) For any sequence of data $t_n(\boldsymbol{y}_n) \to \boldsymbol{t}$, the transformed fiducial distribution measures converge weakly $R_{n,\boldsymbol{y}_n} \Xi_n^{-1} \xrightarrow{\mathcal{W}} R_{\boldsymbol{t}}$.

Theorem 3. Suppose Assumption 1 holds. Fix a desired coverage $0 < \alpha < 1$. For any observed data $\boldsymbol{y}_n$, select an open set $C_n(\boldsymbol{y}_n)$ satisfying: (i) $R_{n,\boldsymbol{y}_n}(C_n(\boldsymbol{y}_n)) = \alpha$; (ii) $t_n(\boldsymbol{y}_n) \to \boldsymbol{t}$ implies $\Xi_n(C_n(\boldsymbol{y}_n)) \to C(\boldsymbol{t})$; (iii) the set $\mathcal{V}_{t_2} = \{(\boldsymbol{v}_1, \boldsymbol{v}_2) : Q_{t_1}(\boldsymbol{v}_1) \in C(\boldsymbol{t}) \text{ and } \boldsymbol{t}_2 = \boldsymbol{H}_2(\boldsymbol{v}_2)\}$ is invariant in $\boldsymbol{t}_1$.

Then the sets $C_n(\boldsymbol{y}_n)$ are α asymptotic confidence sets.

The theorem provides a condition on how various sets of a fixed fiducial probability need to be linked together across different observed datasets to make up a valid confidence set. To understand the key Condition (iii) notice that it assumes the sets $C(\boldsymbol{t})$ are obtained by de-pivoting a common set $\mathcal{V}_{t_2}$. In particular if the limiting data-generating equation $\boldsymbol{T}_1 = \boldsymbol{H}_1(\boldsymbol{V}_1, \boldsymbol{\xi})$ has group structure, Condition (iii) is equivalent to assuming the sets $C(\boldsymbol{t})$ are group invariant in $\boldsymbol{t}_1$. The conditions on the limiting data-generating equation were partially inspired by results for Inferential Models of Martin and Liu (2015a). The proof of Theorem 3 is in Appendix C. Also, this corollary follows immediately:

Corollary 1. Any model that satisfies the assumptions of Theorem 2 satisfies Assumption 1. In particular, for any fixed interior point $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}^0$ the limiting data-generating equation $\boldsymbol{T} = \boldsymbol{\xi} + \boldsymbol{V}$ where the random vector $\boldsymbol{V} \sim N(0, I(\boldsymbol{\theta}_0)^{-1})$. The transformations are $t_n(\boldsymbol{y}_n) = n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$, $\Xi_n(\boldsymbol{\theta}) = n^{1/2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$ and $\boldsymbol{\xi}_0 = 0$. Any collection of sets $C_n(\boldsymbol{y}_n)$ that in the limit becomes location invariant will form asymptotic confidence intervals.

Most of the theoretical results for GFI in the literature were derived in regular statistical problems and are covered by Corollary 1. Notice that in the regular case the limiting data-generating equation has no ancillary part. The next example shows that the ancillary part in Theorem 3 is needed in some nonregular cases.

Example 5 (Example 4 continued). Recall that $Y_i = a(\theta) + b(\theta)U_i$, $i = 1, \ldots, n$, where U_i are iid $U(-1, 1)$. We assume that $a'(\theta) > |b'(\theta)|$ for $\theta \in \boldsymbol{\Theta}$ so that the GFD $R_{n,\boldsymbol{y}_n}$ has a density given by (9). Fix an interior point $\theta_0 \in \boldsymbol{\Theta}^0$. To verify conditions of Theorem 3, we need to define the limiting data-generating equation, and the transformations t_n and Ξ_n. We start with the limiting data-generating process:

$$T_1 = \xi + V_1, \quad T_2 = V_2,$$

where $V_1 = (E_1 - E_2)/2$, $V_2 = (E_1 + E_2)/2$ with E_1, E_2 are independent, $E_1 \sim \exp[\{a'(\theta_0) - b'(\theta_0)\}/\{2b(\theta_0)\}]$ and $E_2 \sim \exp[\{a'(\theta_0) + b'(\theta_0)\}/\{2b(\theta_0)\}]$. The density of the limiting GFD is therefore proportional to

$$r(\xi \mid \boldsymbol{t}) \propto e^{-\xi \{\log b(\theta_0)\}'} I_{(T_1 - T_2, T_1 + T_2)}(\xi).$$

The fact that Assumption 1, Part 2 is satisfied follows immediately.

Next, define the transformations

$$t_n(\boldsymbol{y}) = n \begin{pmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{pmatrix} \cdot \begin{pmatrix} \dfrac{y_{(1)} - (a(\theta_0) - b(\theta_0))}{a'(\theta_0) - b'(\theta_0)} \\ \dfrac{a(\theta_0) + b(\theta_0) - y_{(n)}}{a'(\theta_0) + b'(\theta_0)} \end{pmatrix},$$

$$\Xi_n(\theta) = n(\theta - \theta_0).$$

Simple calculations show that Assumption 1, Part 1 and 3 are satisfied with $\boldsymbol{\xi}_0 = 0$.

Finally, notice that any collection of sets of fiducial probability α that in the limit becomes location invariant in t_1 (such as one sided or equal tailed intervals) are asymptotic α confidence intervals.

2.5. Practical Use of GFI

From a practical point of view, GFI is used in a way similar to the use of a posterior computed using a default (objective) prior, such as probability matching, reference, or flat prior. The main technical difference is that the objective prior is replaced by a data-dependent Jacobian (4). This data dependence can in some examples lead to the existence of second-order matching GFD even when only first-order matching is available with the nondata-dependent priors (Majumder and Hannig 2015). Some argued (Welch and Peers 1963; Martin and Walker 2014) that data-dependent priors are essential in achieving superior frequentist properties in complex statistical problems.

First, we suggest using a set of fiducial probability $1 - \alpha$ and of a good shape (such as one sided or equal tailed) as an approximate $1 - \alpha$ confidence interval, see Theorem 3. Next, the mean or median of the GFD can be used for point estimation.

GFDs can also be used for predicting future observations. This is done by plugging in a random variable having the GFD (2) for the parameter into the data-generating equation for the new observations. This approach produces a predictive distribution that accommodates in a natural way both the uncertainty in the parameter estimation and the randomness of the future data. More details are in Wang, Hannig, and Iyer (2012a).

GFDs are rarely available in closed form. Therefore, we often need to use a Markov Chain Monte Carlo (MCMC) method such as a Metropolis–Hastings or Gibbs sampler to obtain a sample from the GFD. While the basic issues facing implementation of the MCMC procedures are similar for both Bayesian and generalized fiducial problems, there are specific challenges related to generalized fiducial procedures. We discuss some computational issues in Section 5.

3. Model Selection in GFI

Hannig and Lee (2009) introduced model selection into the GFI paradigm in the context of wavelet regression. The presentation here is reexpressed using definition (2). There are two main ingredients needed for an effective fiducial model selection. The first is to include the model as one of the parameters and the second is to include penalization in the data-generating equation.

Consider a finite collection of models $\mathcal{M}$. The data-generating equation is

$$\boldsymbol{Y} = \boldsymbol{G}(M, \boldsymbol{\theta}_M, \boldsymbol{U}), \qquad M \in \mathcal{M},\ \boldsymbol{\theta}_M \in \boldsymbol{\Theta}_M, \tag{11}$$

where $\boldsymbol{Y}$ is the observations, M is the model considered, $\boldsymbol{\theta}_M$ are the parameters associated with model M, and $\boldsymbol{U}$ is a random vector of fully known distribution independent of any parameters.

Denote the number of parameters in the model M by $|M|$. Similar to MLE, an important issue needing to be solved is that GFI tends to favor models with more parameters over ones with fewer parameters. Therefore, an outside penalty accounting for our preference toward parsimony needs to be incorporated in the model. See Appendix D for more details.

In Hannig and Lee (2009), a novel way of adding a penalty into the GFI framework is proposed. In particular, for each model M they proposed augmenting the data-generating Equation (11) by

$$0 = P_k, \quad k = 1, \ldots, \min(|M|, n), \tag{12}$$

where P_k are iid continuous random variables with $f_P(0) = q$ independent of $\boldsymbol{U}$, and q is a constant determined by the penalty. (Based on ideas from the minimum description length principle Hannig and Lee (2009) recommended using $q = n^{-1/2}$ as the default penalty.) Notice that the number of additional equations is the same as the number of unknown parameters in the model.

For the augmented data-generating equation, we have the following theorem. This theorem has never been published before but it does implicitly appear in Hannig and Lee (2009) and Lai, Hannig, and Lee (2015). For completeness, we provide a proof in Appendix D.

Theorem 4. Let us suppose the identifiability Assumption D.1 in Appendix D holds and that each of the models satisfy assumptions of Theorem 1 (in particular $|M| \le n$). Then the marginal generalized fiducial probability of model M is

$$r(M|\boldsymbol{y}) = \frac{q^{|M|} \int_{\boldsymbol{\Theta}_M} f_M(\boldsymbol{y}, \boldsymbol{\theta}_M) J_M(\boldsymbol{y}, \boldsymbol{\theta}_M)\, d\boldsymbol{\theta}_M}{\sum_{M' \in \mathcal{M}} q^{|M'|} \int_{\boldsymbol{\Theta}_{M'}} f_{M'}(\boldsymbol{y}, \boldsymbol{\theta}_{M'}) J_{M'}(\boldsymbol{y}, \boldsymbol{\theta}_{M'})\, d\boldsymbol{\theta}_{M'}}, \tag{13}$$

where $f_M(\boldsymbol{y}, \boldsymbol{\theta}_M)$ is the likelihood and $J_M(\boldsymbol{y}, \boldsymbol{\theta}_M)$ is the Jacobian function computed using (4) for each fixed model M.

Remark 7. The quantity $r(M|\boldsymbol{y})$ can be used for inference in the usual way. For example, *fiducial factor*: the ratio $r(M_1|\boldsymbol{y})/r(M_2|\boldsymbol{y})$, can be used in the same way as a Bayes factor. As discussed in Berger and Pericchi (2001), one of the issues with the use of improper priors in Bayesian model selection is the presence of arbitrary scaling constant. While this is not a problem when a single model is considered, because the arbitrary constant cancels, it becomes a problem for model selection. An advantage of GFD is that the Jacobian function (4) comes with a scaling constant attached to it. In fact, the fiducial factors are closely related to the *intrinsic factors* of Berger and Pericchi (1996, 2001). This can be seen from the fact that for the minimal training sample ($n = |M|$), we usually have $\int_{\boldsymbol{\Theta}_M} f_M(\boldsymbol{y}, \boldsymbol{\theta}_M) J_M(\boldsymbol{y}, \boldsymbol{\theta}_M)\, d\boldsymbol{\theta}_M = 1$.

Similarly, the quantity $r(M|\boldsymbol{y})$ can also be used for fiducial model averaging much akin to the Bayesian model averaging (Hoeting et al. 1999).

We illustrate the use of this model selection on two examples, wavelet regression (Hannig and Lee 2009) and ultra high-dimensional regression (Lai, Hannig, and Lee 2015).

3.1. Wavelet Regression

Suppose n-observed equispaced data points $\{x_i\}_{i=1}^n$ satisfy the following model

$$X_i = g_i + \epsilon_i,$$

where $\boldsymbol{g} = (g_1, \ldots, g_n)^\top$ is the true unknown regression function and ϵ_i's are independent standard normal random variables with mean 0 and variance σ^2, and $n = 2^{J+1}$ is an integer power of 2.

Most wavelet regression methods consist of three steps. The first step is to apply a forward wavelet transform to the data $\boldsymbol{y}$ and obtain the empirical wavelet coefficients $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x}$. Here, $\boldsymbol{H}$ is the discrete wavelet transform matrix. The second step is to apply a shrinkage operation to $\boldsymbol{y}$ to obtain an estimate $\hat{\boldsymbol{d}}$ for the true wavelet coefficients $\boldsymbol{d} = \boldsymbol{H}\boldsymbol{g}$. Lastly, the regression estimate $\hat{\boldsymbol{g}} = (\hat{g}_1, \ldots, \hat{g}_n)^\top$ for $\boldsymbol{g}$ is computed via the inverse discrete wavelet transform: $\hat{\boldsymbol{g}} = \boldsymbol{H}^\top \hat{\boldsymbol{d}}$. The second step of wavelet shrinkage is important because it is the step where statistical estimation is performed. Hannig and Lee (2009) used GFI to perform the second step. Apparently this is the first published work where Fisher's fiducial idea is applied to a nonparametric problem.

Due to the orthonormality of the discrete wavelet transform matrix $\boldsymbol{H}$, a model for the empirical wavelet coefficients is $\boldsymbol{Y} = \boldsymbol{d} + \sigma\boldsymbol{U}$ with $\boldsymbol{U}$ being a n-dimensional vector of independent $N(0, 1)$ random variables. The assumption of sparsity implies that many of the entries in the vector $\boldsymbol{d}$ are zero. This allows us to cast this as a model selection problem, where the model M is the list of nonzero entries. The data-generating Equation (11) becomes

$$Y_k = \begin{cases} d_k + \sigma U_k, & k \in M, \\ \sigma U_k, & k \in M^{\complement}. \end{cases}$$

Notice that $\boldsymbol{\theta}_M = \{\sigma^2, d_k\ k \in M\}$. As discussed above, we augment the data-generating equations by (12) with $q = n^{-1/2}$.

It follows from Theorem 4 that the GFD has generalized density proportional to

$$r(\sigma^2, \boldsymbol{d}, M) \propto (\sigma^{-2})^{\frac{n}{2}+1} \frac{\sum_{j \in M^{\complement}} |y_j|}{n - |M|} \times \exp\left[-\frac{|M| \log n}{2} - \frac{\left\{ \sum_{k \in M} (d_k - y_k)^2 + \sum_{i \in M^{\complement}} y_i^2 \right\}}{2\sigma^2} \right] \times \prod_{i \in M^{\complement}} \delta_0(d_i), \tag{14}$$

where $\delta_0(s)$ is the Dirac function, that is, $\int_A \delta_0(s)\,ds = 1$ if $0 \in A$ and 0 otherwise. The term $1/(n - |M|)$ is an additional normalization term introduced to account for the number of the elements in the sum above it.

The normalizing constant in (14) cannot be computed in a closed form so a sample from $r(\sigma^2, \boldsymbol{d}, I)$ will have to be simulated using MCMC techniques. Note that the GFD is defined in the wavelet domain. Hannig and Lee (2009) used the inverse wavelet transform to define a GFD on the function domain.

Additionally, Hannig and Lee (2009) also assumed that M satisfies a tree condition (Lee 2002). This condition states that if a coefficient is thresholded, all its descendants have to be thresholded too; the exact formulation is in Hannig and Lee (2009). This constraint greatly reduces the search space and allows for both efficient calculations and clean theoretical results. In the article, they reported a simulation study showing small sample performance superior to the alternative methods considered and proved an asymptotic theorem guaranteeing asymptotic consistency of the fiducial model selection.

3.2. Ultra High-Dimensional Regression

Lai, Hannig, and Lee (2015) extended the ideas of fiducial model selection to the ultra high-dimensional regression setting. The most natural data-generating equation for this model is

$$\boldsymbol{Y} = \boldsymbol{G}(M, \boldsymbol{\beta}_M, \sigma^2, \boldsymbol{Z}) = \boldsymbol{X}_M\boldsymbol{\beta}_M + \sigma\boldsymbol{Z},$$

where $\boldsymbol{Y}$ represents the observations, M is the model considered (collection of parameters that are nonzero), $\boldsymbol{X}_M$ is the design matrix for model M, $\boldsymbol{\beta}_M \in \mathbb{R}^{|M|}$ and $\sigma > 0$ are parameters, and $\boldsymbol{Z}$ is a vector of iid standard normal random variables. For computational expediency, they suggested using a sufficient-ancillary transformation that yields the same Jacobian function as the l_2 Jacobian discussed in Section 2.3. The Jacobian function used is

$$J_M(\boldsymbol{y}, \boldsymbol{\theta}_M) = \sigma^{-1}|\det(\boldsymbol{X}_M'\boldsymbol{X}_M)|^{\frac{1}{2}}\mathrm{RSS}_M^{\frac{1}{2}}.$$

The standard Minimum Description Length (MDL) penalty $n^{-|M|/2}$ was not designed to handle ultra high-dimensional problems. Inspired by the Extended Bayesian Information Criterion (EBIC) penalty of Chen and Chen (2008), Lai, Hannig, and Lee (2015) proposed extending the penalty by modifying (12) to

$$0 = B_{|M|}, \quad 0 = P_k,\ k = 1, \ldots, |M|,$$

where $|M|$ is the dimension of M, B_m is a Bernoulli$(1 - r_m)$ random variable that penalizes for the number of models that have the same size m; and P_i are iid continuous random variables with $f_P(0) = q$ independent of B_m that penalize for the size of models. Following the recommendation of Hannig and Lee (2009), we select $q = n^{-1/2}$. Additionally, we select $r_m = \binom{p}{m}^{-\gamma}$, where p is the number of parameters in the full model. The second choice is to penalize for the fact that there is a large number of models that all have the same size. The most natural choice is $\gamma = 1$ for which r_m is the probability of randomly selecting a model M from all models of size m. However, to match the EBIC penalty of Chen and Chen (2008), we allow for other choices of γ.

We assume that for any size m, the residual vectors $\{\boldsymbol{I} - \boldsymbol{X}_M(\boldsymbol{X}_M^\top\boldsymbol{X}_M)^{-1}\boldsymbol{X}_M^\top\}\boldsymbol{y}/\mathrm{RSS}_M$ are distinct for all the models $M \in \mathcal{M}'$ of size m, so that the identifiability Assumption D.1 is satisfied. Theorem 4 implies

$$r(M|\boldsymbol{y}) \propto R_\gamma(M) = \Gamma\left(\frac{n - |M|}{2}\right)(\pi\,\mathrm{RSS}_M)^{-\frac{n-|M|-1}{2}}\, n^{-\frac{|M|+1}{2}}\binom{p}{|M|}^{-\gamma}. \quad (15)$$

Similar to the tree constraint of the previous subsection, Lai, Hannig, and Lee (2015) additionally reduced the number of models by constructing a class of candidate models, denoted as $\mathcal{M}'$. This $\mathcal{M}'$ should satisfy the following two properties: the number of models in $\mathcal{M}'$ is small and it contains the true model and models that have nonnegligible values of $r_\gamma(M)$. To construct $\mathcal{M}'$, they first apply the sure independence screening (SIS) procedure of Fan and Lv (2008) and then apply LASSO and/or SCAD to those p' predictors that survived SIS, and take all those models that lie on the solution path as $\mathcal{M}'$. Note that constructing $\mathcal{M}'$ in this way will ensure the true model is captured in $\mathcal{M}'$ with high probability (Fan and Lv 2008).

Lai, Hannig, and Lee (2015) showed good properties of the GFI solution both by simulation and theoretical considerations. In particular, they proved a consistency theorem provided under the following conditions.

Let M be any model, M_0 be the true model, and $\boldsymbol{H}_M$ be the projection matrix of $\boldsymbol{X}_M$, that is, $\boldsymbol{H}_M = \boldsymbol{X}_M(\boldsymbol{X}_M^\top\boldsymbol{X}_M)^{-1}\boldsymbol{X}_M^\top$. Define $\Delta_M = ||\boldsymbol{\mu} - \boldsymbol{H}_M\boldsymbol{\mu}||^2$, where $\boldsymbol{\mu} = E(\boldsymbol{Y}) = \boldsymbol{X}_{M_0}\boldsymbol{\beta}_{M_0}$. Throughout this subsection, we assume the following identifiability condition holds:

$$\lim_{n\to\infty} \min\left\{\frac{\Delta_M}{|M_0|\log p} :\ M_0 \not\subset M, |M| \le k|M_0|\right\} = \infty \quad (16)$$

for some fixed $k > 1$. Condition (16) is closely related to the sparse Riesz condition (Zhang and Huang 2008).

Let $\mathcal{M}$ be the collection of models such that $\mathcal{M} = \{M :\ |M| \le k|M_0|\}$ for some fixed k. The restriction $|M| \le k|M_0|$ is imposed because in practice we only consider models with size comparable with the true model.

If p is large, a variable screening procedure to reduce the size is still needed. This variable screening procedure should result in a class of candidate models $\mathcal{M}'$, which satisfies

$$P(M_0 \in \mathcal{M}') \to 1 \quad \text{and} \quad \log(m_j') = o(j\log n), \quad (17)$$

where $\mathcal{M}_j'$ contains all models in $\mathcal{M}'$ that are of size j, and m_j' is the number of models in $\mathcal{M}_j'$. The first condition in (17) guarantees the model class contains the true model, at least asymptotically. The second condition in (17) ensures that the size of the model class is not too large. The authors report small sample performance preferable to competing methods as determined by simulation study and prove asymptotic consistency of the fiducial model selection algorithm.

4. GFI for Discrete and Interval Data

Most of the material presented in Sections 2 and 3 was developed for exactly observed continuous distributions. This section discusses discrete and discretized observations.

When the observations are discrete then there is no problem with the Borel paradox and the limiting distribution in (2) can be easily computed; see Remark 4. In particular, if we define $Q_{\boldsymbol{y}}(\boldsymbol{u}) = \{\boldsymbol{\theta} : \boldsymbol{y} = \boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta})\}$ the GFD is the conditional distribution

$$V[Q_{\boldsymbol{y}}(\boldsymbol{U}^\star)] \mid \{Q_{\boldsymbol{y}}(\boldsymbol{U}^\star) \neq \emptyset\}, \tag{18}$$

where $V[A]$ selects a (possibly random) element of the closure of the set A and $\boldsymbol{U}^\star$ is an independent copy of $\boldsymbol{U}$. If $A = (a, b)$ is a finite interval, then we recommend a rule that selects one of the endpoints a or b at random independent of $\boldsymbol{U}^\star$ (Hannig 2009). This selection maximizes the variance of the GFD, has been also called "half correction" (Efron 1998; Schweder and Hjort 2002; Hannig and Xie 2012) and is closely related to the well-known continuity correction used in normal approximations.

4.1. Some Common Discrete Distributions

In this subsection, we compute the GFDs for parameters of several popular discrete distributions.

Example 6. Let X be a random variable with distribution function $F(y|\theta)$. Assume there is $\mathcal{Y}$ so that $P_\theta(Y \in \mathcal{Y}) = 1$ for all θ, and for each fixed $y \in \mathcal{Y}$ the distribution function is either a nonincreasing function of θ, spanning the whole interval $(0, 1)$, or a constant equal to 1. Similarly, the left limit $F(y_-|\theta)$ is also either a nonincreasing function of θ spanning the whole interval $(0, 1)$, or a constant equal to 0.

Define the near inverse $F^-(a|\theta) = \inf\{y : F(y|\theta) \geq a\}$. It is well known (Casella and Berger 2002) that if $U \sim \mathrm{U}(0,1)$, $Y = F^-(U|\theta)$ has the correct distribution and we use this association as a data-generating equation.

Next, it follows that both $Q_y^+(u) = \sup\{\theta : F(y|\theta) = u\}$ and $Q_y^-(u) = \inf\{\theta : F(y_-|\theta) = u\}$ exist and satisfy $F(y|Q_y^+(u)) = u$ and $F(y_-|Q_y^-(u)) = u$. Consequently if $\boldsymbol{U}^\star$ is an independent copy of $\boldsymbol{U}$

$$P(Q_y^+(u) \leq t) = 1 - F(y|t) \quad \text{and} \quad P(Q_y^-(u) \leq t) = 1 - F(y_-|t).$$

Finally, notice that for all $u \in (0, 1)$ the function $F^-(u|\theta)$ is nondecreasing in θ and the closure of the inverse image $\bar{Q}_y(u) = \{Q_y^-(u), Q_y^+(u)\}$. Since the condition in (18) has probability 1, there is no conditioning and the half corrected GFD has distribution function

$$R(\theta|y) = 1 - \frac{F(y|\theta) + F(y_-|\theta)}{2}.$$

If either of the distribution function is constant, we interpret it as a point mass at the appropriate boundary of the parameter space.

Analogous argument shows that if the distribution function and its left limit were increasing in θ than the half corrected GFD would have distribution function

$$R(\theta|y) = \frac{F(y|\theta) + F(y_-|\theta)}{2}.$$

Using this result, we provide a list of the half corrected GFDs for three well-known discrete distributions. Here, we understand $\mathrm{Beta}(0, n + 1)$ and $\mathrm{Beta}(x + 1, 0)$ as the degenerate distributions (Dirac measure) on 0 and 1, respectively. Similarly, we understand $\Gamma(0, 1)$ as the degenerate distribution (Dirac measure) on 0.

- $X \sim \mathrm{Binomial}(n, p)$ with n known. GFD is the 50–50 mixture of $\mathrm{Beta}(x + 1, n - x)$ and $\mathrm{Beta}(x, n - x + 1)$ distributions, see Hannig (2009).
- $X \sim$ Poisson(λ). GFD is the 50–50 mixture of $\mathrm{Gamma}(x + 1, 1)$ and $\mathrm{Gamma}(x, 1)$ distributions, see Dempster (2008).
- $X \sim$ Negative Binomial(r, p) with r known. GFD is the 50–50 mixture of $\mathrm{Beta}(r, x - r + 1)$ and $\mathrm{Beta}(r, x - r)$ distributions, see Hannig (2014).

Example 7. Next we consider $Y \sim \mathrm{Multinomial}(n, p_1, \ldots, p_k)$, where n is known and $p_i \geq 0$, $\sum_{i=1}^k p_i = 1$ are unknown.

When the categories of the multinomial have a natural ordering, Hannig (2009) suggested to write $\boldsymbol{Y} = \sum_{i=1}^n \boldsymbol{X}_i$, $q_l = \sum_{i=1}^l p_i$ and model each X_i through the data-generating equation

$$\boldsymbol{X}_i = \left(I_{(0,q_1)}(U_i), I_{[q_1,q_2)}(U_i), \ldots, I_{[q_{k-1},1)}(U_i)\right)^\top, \quad i = 1, \ldots, n,$$

where $U_1, \ldots U_n$ are iid $\mathrm{U}(0, 1)$ random variables. Denote the first quadrant $\mathcal{Q} = \{\boldsymbol{q} : 0 \leq q_1 \leq \ldots q_{k-1} \leq 1\}$. Hannig (2009) showed that the GFD (18) for $\boldsymbol{q}$ is given by

$$V[\{\boldsymbol{q}^\star \in \mathcal{Q} : U^\star_{(\sum_{j=1}^i y_j)} \leq q_i^\star \leq U^\star_{(1+\sum_{j=1}^i y_j)}, \quad i = 1 \ldots, k-1\}],$$

where y_i is the ith component of the observed $\boldsymbol{y}$ and $U^\star_{(j)}$ is the jth order statistics of $U_1^\star, \ldots, U_n^\star$, which is an independent copy of $\boldsymbol{U}$. The GFD for $\boldsymbol{p}$ is then obtained by a simple transformation. Hannig (2009) showed good asymptotic and small sample properties of this GFD.

A drawback of the solution above is its dependency on the ordering of the categories. Lawrence et al. (2009) provided a solution that does not rely on a potentially arbitrary ordering of the categories. Their approach starts from analyzing each coordinate of $\boldsymbol{Y}$ individually.

As can be seen in Example 6, the fiducial inversion of each coordinate when ignoring the others gives a relationship $U_i \leq p_i \leq 1$ where $U_i \sim \mathrm{Beta}(y_i, 1)$ are independent. Additionally, the fact that $\sum_{i=1}^k p_i = 1$ imposes a condition $\sum_{i=1}^k U_i \leq 1$. Consider the following random vector with its distribution taken as the conditional distribution

$$(W_0^\star, W_1^\star, \ldots, W_k^\star) \sim (1 - U_1 - \cdots - U_k, U_1, \ldots, U_k) \mid \{U_1 + \cdots + U_k \leq 1\}.$$

A straightforward calculation shows that the vector $\boldsymbol{W}$ follows Dirichlet$(1, y_1, \ldots, y_k)$ distribution. Writing $Q_{\boldsymbol{y}}(\boldsymbol{w}) = \{\boldsymbol{p} : w_i \leq p_i,\ i = 1, \ldots, k\}$ the GFD is $V[Q_{\boldsymbol{y}}(\boldsymbol{W}^\star)]$.

Denote by e_i, $i = 1, \ldots k$ the coordinate unit vectors in $\mathbb{R}^k$. Notice that the set $Q_{\boldsymbol{y}}(\boldsymbol{w})$ is a simplex with vertexes $\{(w_1, \ldots, w_k) + e_i w_0,\ i = 1, \ldots, k\}$. The selection rule V analogous to the half correction selects each vertex with equal probability and the GFD is an equal probability $(1/k)$ mixture of Dirichlet$(Y_1 + 1, Y_2, \ldots, Y_k)$, ..., Dirichlet$(Y_1, Y_2, \ldots, Y_k + 1)$.

4.2. Median Lethal Dose (LD50)

Consider an experiment involving k dose levels $x_1, x_2, \ldots, x_k$. Each dose level x_i is administered to n_i subjects with y_i positive responses, $i = 1, 2, \ldots, k$. Assume that the relationship between

dose level x_i and the probability p_i of a positive response can be represented by the logistic-linear model, given by

$$\text{logit}(p_i) = \beta_1 x_i + \beta_0 = \beta_1(x_i - \mu),$$

where $\mu = -\beta_0/\beta_1$ represents the median lethal dose (LD50) and $\text{logit}(p_i) = \log\{p_i/(1-p_i)\}$. The parameter of interest LD50 is frequently of interest in many applied fields. Examples include a measure toxicity of a compound in a species in quantal bioassay experiments and measure of difficulty in item response models.

There are three classical methods for estimating LD50: the delta method, Fieller's method, and the likelihood ratio method. If the dose–response curve is steep relative to the spread of doses, then there may be no dose groups, or at most one dose group, with observed mortalities strictly between 0% and 100%. In such cases, the maximum likelihood estimator of β_1 is not calculable and the Delta method and Fieller's method fail to provide a confidence set. Furthermore, when the standard Wald test does not reject the null hypothesis $\beta_1 = 0$, Fieller's confidence sets are either the entire real line or unions of disjoint intervals. Likewise, if the null hypothesis could not be rejected by the likelihood ratio test, the likelihood ratio confidence sets are either the entire real line or unions of disjoint intervals.

E, Hannig, and Iyer (2009) proposed a generalized fiducial solution that does not suffer from these issues. They based their inference on the following data-generating equation: Let Y_{ij}, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$ denote the jth subject's response to the dose level x_i. Since Y_{ij} follows a Bernoulli distribution with success probability $p_i = \text{antilogit}(\beta_0 + \beta_1 x_i)$:

$$Y_{ij} = I_{(0,\text{antilogit}(\beta_0+\beta_1 x_i))}(U_{ij}), \quad j = 1, \ldots, n_i, \ i = 1, \ldots, k.$$

Here (β_0, β_1) are unknown parameters and U_{ij} are independent standard uniform random variables.

The GFD is well-defined using (18) and E, Hannig, and Iyer (2009) proposed to use a Gibbs sampler to implement it. They performed a thorough simulation study showing that the generalized fiducial method compares favorably to the classical methods in terms of coverage and median length of the confidence interval for LD(50). Moreover, the generalized fiducial method performed well even in the situation when the classical methods fail. They also proved that the fiducial CIs give asymptotically correct coverage, and that the effect of discretization is negligible in the limit.

4.3. Discretized Observations

In practice, most datasets are rounded off in some manner, say, by a measuring instrument or by storage on a computer. Mathematically speaking, we do not know the exact realized value $\boldsymbol{Y} = \boldsymbol{y}$. Instead we only observe an occurrence of an event $\{\boldsymbol{Y} \in A_{\boldsymbol{y}}\}$, for some multivariate interval $A_{\boldsymbol{y}} = [\boldsymbol{a}, \boldsymbol{b})$ containing $\boldsymbol{y}$ and satisfying $P_{\theta_0}(\boldsymbol{Y}^\star \in A_{\boldsymbol{y}}) > 0$, where $\boldsymbol{Y}^\star = \boldsymbol{G}(\boldsymbol{U}^\star, \boldsymbol{\theta}_0)$ is an independent copy of $\boldsymbol{Y}$.

For example, if the exact value of the random vector $\boldsymbol{Y}$ was $\boldsymbol{y} = (\pi, e, 1.28)$ and due to instrument precision all the values were rounded to one decimal place, our observation would be the event $A_{\boldsymbol{y}} = [3.1, 3.2) \times [2.7, 2.8) \times [1.2, 1.3)$.

Since $P_{\theta_0}(\boldsymbol{Y}^\star \in A_{\boldsymbol{y}}) > 0$, the arguments in Remark 4 still apply and the formula (18) remains valid with $Q_{\boldsymbol{y}}(\boldsymbol{u}) = \{\boldsymbol{\theta} : \boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta}) \in \bar{A}_{\boldsymbol{y}}\}$, where $\bar{A}_{\boldsymbol{y}}$ is the closure of $A_{\boldsymbol{y}}$.

Hannig (2013) proved fiducial Bernstein–von Mises theorem for discretized data. He assumed that we observed discretized iid observations with a distribution function $F(y|\boldsymbol{\theta})$. He set $F^-(a|\boldsymbol{\theta}) = \inf\{y : F(y|\boldsymbol{\theta}) \geq a\}$ and assumed the data-generating equation

$$Y_i = F^-(U_i \mid \boldsymbol{\theta}), \quad i = 1, \ldots, n,$$

where Y_i are random variables, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is a p-dimensional parameter, U_i are iid $U(0, 1)$.

We restate the main theorem in Hannig (2013) in the language of this review article:

Theorem 5 (Hannig 2013). Suppose Assumption E.1 in Appendix E holds. Then the GFD defined by (18) has the same asymptotically normal distribution and satisfies Assumption 1 regardless the choice of $V[.]$. Consequently, any collection of sets $C_n(\boldsymbol{y}_n)$ that in the limit becomes location invariant will form asymptotically correct confidence intervals.

4.4. Linear Mixed Models

Despite the long history of inference procedures for normal linear mixed models, a well-performing, unified inference method is lacking. Analysis of variance (ANOVA)-based methods offer, what tends to be, model-specific solutions. Bayesian methods allow for solutions to very complex models, but determining an appropriate prior distribution can be confusing.

Cisewski and Hannig (2012) proposed the use of GFI for discretized linear mixed models that avoids the issues mentioned above. They started with the following data-generating equation:

$$\mathbf{Y} = \mathbf{X}\beta + \sum_{i=1}^{r} \sigma_i \sum_{j=1}^{l_i} \boldsymbol{V}_{i,j} U_{i,j},$$

where $\mathbf{X}$ is a known $n \times p$ fixed-effects design matrix, β is the $p \times 1$ vector of fixed effects, $\boldsymbol{V}_{i,j}$ is the $n \times 1$ design vector for level j of random effect i, l_i is the number of levels per random effect i, σ_i^2 is the variance of random effect i, and the $U_{i,j}$ are independent and identically distributed standard normal random variables.

To compute the GFD in (18), Cisewski and Hannig (2012) designed a computationally efficient modification of sequential Monte Carlo (SMC) algorithm (Doucet, De Freitas, and Gordon 2001; Del Moral, Doucet, and Jasra 2006; Douc and Moulines 2008). The fiducial implementation includes a custom design resampling and modification step that greatly improves the efficiency of the SMC algorithm for this model.

Cisewski and Hannig (2012) performed a thorough simulation study showing that the proposed method yields confidence interval estimation for all parameters of balanced and unbalanced normal linear mixed models. The fiducial intervals were as good as or better than the best tailor made ANOVA-based solutions for the simulation scenarios covered. In addition, for the models considered by Cisewski and Hannig (2012)

and for the prior selected based on recommendations in the literature, the Bayesian interval lengths were not generally competitive with the other methods used in the study.

The authors point out that even though more variation was incorporated into the data for the generalized fiducial method due to the use of discretized data, the generalized fiducial method tended to maintain stated coverage (or be conservative) while having average interval lengths comparable or shorter than other methods even though the competing methods assumed the data are observed exactly.

5. Computational Issues

This section presents some computational challenges involved when applying GFI in practice and some possible solutions to solve these challenges.

For any given model, we recall that the GFD is defined as the weak limit in (2) and under fairly general conditions, the weak limit has a density $r(\boldsymbol{\theta}|\boldsymbol{y})$ given in (3). This density can often be used directly to form estimates and asymptotic confidence intervals for the model parameters, in a similar manner as the density of the posterior distribution in the Bayesian paradigm. Standard sampling techniques such as MCMC, importance sampling, or sequential Monte Carlo have been successfully implemented, for example, Hannig et al. (2006a), Hannig (2009), Hannig and Lee (2009), Wandler and Hannig (2012b), and Cisewski and Hannig (2012).

The exact form of generalized fiducial density could be hard to compute. For this reason, Hannig, Lai, and Lee (2014) presented a computationally tractable solution for conducting generalized fiducial inference without knowing the exact closed form of the generalized fiducial distribution.

5.1. Evaluating the Generalized Fiducial Density via Subsampling

In some situations, even the denominator of the density $r(\boldsymbol{\theta}|\boldsymbol{y})$ becomes too complicated to evaluate directly, particularly so when the l_∞ norm is used in (2). In such situations, the function $D(\cdot)$ in (4) is a sum over all possible tuples of length p, that is, $D(A) = \sum_{\boldsymbol{i}=(i_1,\ldots,i_p)} |\det(A)_{\boldsymbol{i}}|$. If we have n observations, there are in total $\binom{n}{p}$ number of possible tuples. If the sum cannot be simplified analytically, one is obliged to compute all $\binom{n}{p}$ terms. Such computations can become prohibitively expensive even for moderate n and p. Appropriate approximations are required to evaluate the density efficiently.

If the observations are iid and l_∞ norm is used, $D(\frac{d}{d\boldsymbol{\theta}}\boldsymbol{G}(\boldsymbol{u},\boldsymbol{\theta})|_{\boldsymbol{u}=\boldsymbol{G}^{-1}(\boldsymbol{y},\boldsymbol{\theta})})$ is a U-statistic. Given the strong dependency of the terms in $D(\cdot)$, it seems possible to use much less than $\binom{n}{p}$ terms for approximation without loss of accuracy. Blom (1976) showed that incomplete U-statistic based on random selection of K subsamples behaves very similar to the complete U-statistic when n and K are large. On the basis of this result, Hannig (2009), and its follow-up articles, we suggest to replace $D(\cdot)$ by

$$\hat{D}(A;\mathcal{I}_K) = \sum_{\boldsymbol{i}\in\mathcal{I}_K} |\det(A)_{\boldsymbol{i}}|\,,$$

where $\mathcal{I}_K$ is a random selection of K different p-tuples. Numerical simulations confirm that this approximation is very promising for a wide range of applications. In practice, a common choice of K would be in the order of hundreds. One may want to choose K keeping in mind that a small K may fail to yield a good enough approximation. On the other hand, a large value of K would cause too much computations and it may be not favorable.

In most algorithms such as an MCMC sampler, the density is repeatedly evaluated for different values of $\boldsymbol{\theta}$. We recommend to keep the same choice of $\mathcal{I}_K$ for different values of $\boldsymbol{\theta}$ to gain stability of the algorithm.

The above discussion also applies to the generalized fiducial density (13) when model selection is involved.

6. Concluding Remarks and Open Problems

After many years of investigations, the authors and collaborators have demonstrated that GFI is a useful and promising approach for conducting statistical inference. GFI has been validated by asymptotic theory and by simulation in numerous small sample problems. In this article, we have summarized the latest theoretical and methodological developments and applications of GFI. To conclude, we list some open and important research problems about GFI.

1. As mentioned earlier, the choice of data-generating equation $\boldsymbol{G}$ in (1) is not unique for many problems. Based on our practical experience gained from simulations, GFD-based intervals are usually conservative and often quite short as compared to competing methods for small sample sizes. This property is not well understood as traditional asymptotic tools (including higher order asymptotics) do not explain it. Understanding this nonasymptotic phenomenon will likely help both with deeper understanding of GFI and the optimal choice of $\boldsymbol{G}$. Although our numerical experience suggests that different choices of $\boldsymbol{G}$ only lead to small differences in practical performances, it would still be important to develop an objective method for choosing $\boldsymbol{G}$.
2. As an interesting alternative, one could modify the GFD definition (2) by adding a penalty term $p(\cdot)$ on $\boldsymbol{\theta}$ to encourage sparse solutions:

$$\lim_{\epsilon\to 0}\Bigg[\arg\min_{\boldsymbol{\theta}^\star} \|\boldsymbol{y}-\boldsymbol{G}(\boldsymbol{U}^\star,\boldsymbol{\theta}^\star)\| + p(\boldsymbol{\theta}^\star)\,\Big|\,\|\boldsymbol{y}-\boldsymbol{G}(\boldsymbol{U}^\star,\boldsymbol{\theta}^\star)\|\le\epsilon\Bigg]. \qquad (19)$$

For example, in the context of linear regression with an l_1 penalty $p(\cdot)$, just as the lasso (Tibshirani 1996) and Dantzig selector (Candes and Tao 2007) do, (19) will lead to sparse solutions. We stress that while obtaining *sparse point* estimators through a minimization problem has become a standard technique, (19) produces *sparse distributions* on the parameter space also as a result of optimization. This is different from sparse posterior distributions obtained as a result of sparsity priors. The hope

is that this approach will lead to computationally efficient ways of quantifying uncertainty in model selection procedures.

3. One possible way to gain a deeper philosophical understanding of GFI is to find a general set of conditions under which GFI is in some sense an optimal data-dependent distribution on the parameter space (assuming such a set exists). The work of Taraldsen and Lindqvist (2013) that provides an initial result on a connection between decision theory and fiducial inference would be a good starting point.
4. It would be interesting to investigate the performance of GFI when the data-generating equation is misspecified. For example, what would happen to the empirical confidence interval coverages if $N(0, 1)$ is used as the random component when the truth is in fact t with 3 degrees of freedom?

Lastly, we hope that our contributions to GFI will stimulate the growth, usage, and interest of this exciting approach for statistical inference in various research and application communities.

Supplementary Materials

The online supplementary materials contain the appendices for the article, and code for many of the methods in this review.

Acknowledgment

The authors are thankful to Yifan Cui who has found numerous typos in an earlier version of this article. They are also most grateful to the reviewers and the associate editor for their constructive and insightful comments and suggestions.

Funding

Hannig was supported in part by the National Science Foundation under Grant Nos. 1016441, 1633074, and 1512893. Lee was supported in part by the National Science Foundation under Grant Nos. 1209226, 1209232, and 1512945.

References

Barnard, G. A. (1995), "Pivotal Models and the Fiducial Argument," *International Statistical Reviews*, 63, 309–323. [1346]

Bayarri, M. J., Berger, J. O., Forte, A., and García-Donato, G. (2012), "Criteria for Bayesian Model Choice With Application to Variable Selection," *The Annals of Statistics*, 40, 1550–1577. [1347]

Beaumont, M. A., Zhang, W., and Balding, D. J. (2002), "Approximate Bayesian Computation in Population Genetics," *Genetics*, 162, 2025–2035. [1349]

Berger, J. O. (1992), "On the Development of Reference Priors" (with discussion), *Bayesian Statistics*, 4, 35–60. [1347]

Berger, J. O., Bernardo, J. M., and Sun, D. (2009), "The Formal Definition of Reference Priors," *The Annals of Statistics*, 37, 905–938. [1347,1350,1351]

——— (2012), "Objective Priors for Discrete Parameter Spaces," *Journal of the American Statistical Association*, 107, 636–648. [1347]

Berger, J. O., and Pericchi, L. R. (1996), "The Intrinsic Bayes Factor for Model Selection and Prediction," *Journal of the American Statistical Association*, 91, 109–122. [1354]

——— (2001), "Objective Bayesian Methods for Model Selection: Introduction and Comparison," in *Model Selection (IMS Lecture Notes Monogr. Ser., Vol. 38)*, ed. P. Lahiri, Beachwood, OH: Inst. Math. Statist., pp. 135–207. [1354]

Berger, J. O., and Sun, D. (2008), "Objective Priors for the Bivariate Normal Model," *The Annals of Statistics*, 36, 963–982. [1347]

Birnbaum, A. (1961), "On the Foundations of Statistical Inference: Binary Experiments," *The Annals of Mathematical Statistics*, 32, 414–435. [1348]

——— (1962), "On the Foundations of Statistical Inference," *Journal of the American Statistical Association*, 57, 269–326. [1350]

Blom, G. (1976), "Some Properties of Incomplete U-Statistics," *Biometrika*, 63, 573–580. [1358]

Candes, E., and Tao, T. (2007), "The Dantzig Selector: Statistical Estimation When p is Much Larger Than n," *The Annals of Statistics*, 35, 2313–2351. [1358]

Casella, G., and Berger, R. L. (2002), *Statistical Inference* (2nd ed.), Pacific Grove, CA: Wadsworth and Brooks/Cole Advanced Books and Software. [1348,1356]

Chen, J., and Chen, Z. (2008), "Extended Bayesian Information Criteria for Model Selection With Large Model Spaces," *Biometrika*, 95, 759–771. [1355]

Chiang, A. K. L. (2001), "A Simple General Method for Constructing Confidence Intervals for Functions of Variance Components," *Technometrics*, 43, 356–367. [1347]

Cisewski, J., and Hannig, J. (2012), "Generalized Fiducial Inference for Normal Linear Mixed Models," *The Annals of Statistics*, 40, 2102–2127. [1347,1348,1357,1358]

Dawid, A. P., and Stone, M. (1982), "The Functional-Model Basis of Fiducial Inference," *The Annals of Statistics*, 10, 1054–1074. [1346]

Dawid, A. P., Stone, M., and Zidek, J. V. (1973), "Marginalization Paradoxes in Bayesian and Structural Inference," *Journal of the Royal Statistical Society*, Series B, 35, 189–233. [1346]

Del Moral, P., Doucet, A., and Jasra, A. (2006), "Sequential Monte Carlo Samplers," *Journal of the Royal Statistical Society*, Series B, 68, 411–436. [1357]

Dempster, A. P. (1966), "New Methods for Reasoning Towards Posterior Distributions Based on Sample Data," *The Annals of Mathematical Statistics*, 37, 355–374. [1346]

——— (1968), "A Generalization of Bayesian Inference" (with discussion), *Journal of the Royal Statistical Society*, Series B, 30, 205–247. [1346]

——— (2008), "The Dempster-Shafer Calculus for Statisticians," *International Journal of Approximate Reasoning*, 48, 365–377. [1346,1348]

Douc, R., and Moulines, E. (2008), "Limit Theorems for Weighted Samples With Applications to Sequential Monte Carlo Methods," *The Annals of Statistics*, 36, 2344–2376. [1357]

Doucet, A., De Freitas, N., and Gordon, N. (2001), *Sequential Monte Carlo Methods in Practice*, New York: Springer. [1357]

E, L., Hannig, J., and Iyer, H. K. (2008), "Fiducial Intervals for Variance Components in an Unbalanced Two-Component Normal Mixed Linear Model," *Journal of the American Statistical Association*, 103, 854–865. [1347]

——— (2009), "Applications of Generalized Fiducial Inference," Ph.D. Dissertation, Colorado State University, Fort Collins, CO. [1348,1357]

Edlefsen, P. T., Liu, C., and Dempster, A. P. (2009), "Estimating Limits From Poisson Counting Data Using Dempster–Shafer Analysis," *The Annals of Applied Statistics*, 3, 764–790. [1346]

Efron, B. (1998), "R.A. Fisher in the 21st Century," *Statistical Science*, 13, 95–122. [1356]

Fan, J., and Lv, J. (2008), "Sure Independence Screening for Ultrahigh Dimensional Feature Space," *Journal of the Royal Statistical Society*, Series B, 70, 849–911. [1355]

Fisher, R. A. (1922), "On the Mathematical Foundations of Theoretical Statistics," *Philosophical Transactions of the Royal Society of London*, Series A, 222, 309–368. [1346]

——— (1925), "Theory of Statistical Estimation," *Proceedings of the Cambridge Philosophical Society*, 22, 700–725. [1346]

——— (1930), "Inverse Probability," *Proceedings of the Cambridge Philosophical Society*, xxvi, 528–535. [1346]

——— (1933), "The Concepts of Inverse Probability and Fiducial Probability Referring to Unknown Parameters," *Proceedings of the Royal Society of London*, Series A, 139, 343–348. [1346]

——— (1935), "The Fiducial Argument in Statistical Inference," *The Annals of Eugenics*, VI, 91–98. [1346]
Fraser, A. M., Fraser, D. A. S., and Staicu, A.-M. (2009), "The Second Order Ancillary: A Differential View With Continuity," *Bernoulli. Official Journal of the Bernoulli Society for Mathematical Statistics and Probability*, 16, 1208–1223. [1347]
Fraser, D., and Naderi, A. (2008), "Exponential Models: Approximations for Probabilities," *Biometrika*, 94, 1–9. [1347]
Fraser, D., Reid, N., and Wong, A. (2005), "What a Model With Data Says About Theta," *International Journal of Statistical Science*, 3, 163–178. [1347]
Fraser, D. A. S. (1961a), "On Fiducial Inference," *The Annals of Mathematical Statistics*, 32, 661–676. [1346]
——— (1961b), "The Fiducial Method and Invariance," *Biometrika*, 48, 261–280. [1346]
——— (1966), "Structural Probability and a Generalization," *Biometrika*, 53, 1–9. [1346]
——— (1968), *The Structure of Inference*, New York-London-Sydney: Wiley. [1346]
——— (2004), "Ancillaries and Conditional Inference," *Statistical Science*, 19, 333–369. [1347]
——— (2011), "Is Bayes Posterior Just Quick and Dirty Confidence?" *Statistical Science*, 26, 299–316. [1347]
Fraser, D. A. S., Reid, N., Marras, E., and Yi, G. Y. (2010), "Default Priors for Bayesian and Frequentist Inference," *Journal of the Royal Statistical Society*, Series B, 72. [1347,1350]
Glagovskiy, Y. S. (2006), "Construction of Fiducial Confidence Intervals For the Mixture of Cauchy and Normal Distributions," Master's Thesis, Department of Statistics, Colorado State University, Fort Collins, CO. [1347]
Hannig, J. (2009), "On Generalized Fiducial Inference," *Statistica Sinica*, 19, 491–544. [1347,1348,1349,1352,1356,1358]
——— (2013), "Generalized Fiducial Inference via Discretization," *Statistica Sinica*, 23, 489–514. [1347,1348,1349,1350,1357]
——— (2014), Discussion of "On the Birnbaum Argument for the Strong Likelihood Principle" by D. G. Mayo, *Statistical Science*, 29, 254–258. [1358]
Hannig, J., E, L., Abdel-Karim, A., and Iyer, H. K. (2006a), "Simultaneous Fiducial Generalized Confidence Intervals for Ratios of Means of Lognormal Distributions," *Austrian Journal of Statistics*, 35, 261–269. [1347,1358]
Hannig, J., Iyer, H. K., and Patterson, P. (2006b), "Fiducial Generalized Confidence Intervals," *Journal of American Statistical Association*, 101, 254–269. [1347]
Hannig, J., Iyer, H. K., and Wang, J. C.-M. (2007), "Fiducial Approach to Uncertainty Assessment: Accounting for Error Due to Instrument Resolution," *Metrologia*, 44, 476–483. [1347]
Hannig, J., Lai, R. C. S., and Lee, T. C. M. (2014), "Computational Issues of Generalized Fiducial Inference," *Computational Statistics and Data Analysis*, 71, 849–858. [1358]
Hannig, J., and Lee, T. C. M. (2009), "Generalized Fiducial Inference for Wavelet Regression," *Biometrika*, 96, 847–860. [1347,1354,1355,1358]
Hannig, J., Wang, C. M., and Iyer, H. K. (2003), "Uncertainty Calculation for the Ratio of Dependent Measurements," *Metrologia*, 4, 177–186. [1347]
Hannig, J., and Xie, M. (2012), "A Note on Dempster-Shafer Recombinations of Confidence Distributions," *Electrical Journal of Statistics*, 6, 1943–1966. [1347,1356]
Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999), "Bayesian Model Averaging: A Tutorial" (with discussion), *Statistical Science*, 14, 382–417; corrected version available at *http://www.stat.washington.edu/www/research/online/hoetingl999.pdf*. [1354]
Iyer, H. K., and Patterson, P. (2002), "A Recipe for Constructing Generalized Pivotal Quantities and Generalized Confidence Intervals," Technical Report 10, Department of Statistics, Colorado State University, Fort Collins CO. [1350]
Iyer, H. K., Wang, C. M. J., and Mathew, T. (2004), "Models and Confidence Intervals for True Values in Interlaboratory Trials," *Journal of the American Statistical Association*, 99, 1060–1071. [1347]
Jeffreys, H. (1940), "Note on the Behrens-Fisher Formula," *The Annals of Eugenics*, 10, 48–51. [1346]
Lai, R. C. S., Hannig, J., and Lee, T. C. M. (2015), "Generalized Fiducial Inference for Ultra-High Dimensional Regression," *Journal of American Statistical Association*, 110, 760–772. [1347,1354,1355]
Lawrence, E., Liu, C., Vander Wiel, S., and Zhang, J. (2009), "A New Method for Multinomial Inference Using Dempster-Shafer Theory." [1356]
Lee, T. C. M. (2002), "Tree-Based Wavelet Regression for Correlated Data using the Minimum Description Length Principle," *Australian and New Zealand Journal of Statistics*, 44, 23–39. [1355]
Lindley, D. V. (1958), "Fiducial Distributions and Bayes' Theorem," *Journal of the Royal Statistical Society*, Series B, 20, 102–107. [1346]
Liu, Y., and Hannig, J. (2016), "Generalized Fiducial Inference for Binary Logistic Item Response Models," *Psychometrica*, 81, 290–324. [1347]
Majumder, P. A., and Hannig, J. (2015), "Higher Order Asymptotics for Generalized Fiducial Inference," unpublished manuscript. [1347,1353]
Martin, R., and Liu, C. (2013), "Inferential Models: A Framework for Prior-Free Posterior Probabilistic Inference," *Journal of the American Statistical Association*, 108, 301–313. [1346]
——— (2015a), "Conditional Inferential Models: Combining Information for Prior-Free Probabilistic Inference," *Journal of the Royal Statistical Society*, Series B, 77, 195–217. [1346,1353]
——— (2015b), *Inferential Models: Reasoning with Uncertainty* (Vol. 145), Boca Raton, FL: CRC Press. [1346]
——— (2015c), "Marginal Inferential Models: Prior-Free Probabilistic Inference on Interest Parameters," *Journal of the American Statistical Association*, 110, 1621–1631. [1346]
Martin, R., and Walker, S. G. (2014), "Asymptotically Minimax Empirical Bayes Estimation of a Sparse Normal Mean Vector," *Electronic Journal of Statistics*, 8, 2188–2206. [1353]
Martin, R., Zhang, J., and Liu, C. (2010), "Dempster-Shafer Theory and Statistical Inference With Weak Beliefs," *Statistical Science*, 25, 72–87. [1346]
McNally, R. J., Iyer, H. K., and Mathew, T. (2003), "Tests for Individual and Population Bioequivalence Based on Generalized p-Values," *Statistics in Medicine*, 22, 31–53. [1347]
Patterson, P., Hannig, J., and Iyer, H. K. (2004), "Fiducial Generalized Confidence Intervals for Proportionof Conformance," Technical Report 2004/11, Colorado State University, Fort Collins, CO. [1347]
Salome, D. (1998), "Statistical Inference via Fiducial Methods," Ph.D. dissertation, University of Groningen, Groningen, The Netherlands. [1346]
Schweder, T., and Hjort, N. L. (2002), "Confidence and Likelihood," *Scandinavian Journal of Statistics*, 29, 309–332. [1346,1356]
Singh, K., Xie, M., and Strawderman, W. E. (2005), "Combining Information From Independent Sources Through Confidence Distributions," *The Annals of Statistics*, 33, 159–183. [1346]
Sonderegger, D., and Hannig, J. (2014), "Fiducial Theory for Free-Knot Splines," in *Contemporary Developments in Statistical Theory, a Festschrift in honor of Professor Hira L. Koul*, eds. S. Lahiri, A. Schick, A. SenGupta, T. N. Sriram, New York: Springer, pp. 155–189. [1347,1352]
Stevens, W. L. (1950), "Fiducial Limits of the Parameter of a Discontinuous Distribution," *Biometrika*, 37, 117–129. [1346]
Taraldsen, G., and Lindqvist, B. H. (2013), "Fiducial Theory and Optimal Inference," *The Annals of Statistics*, 41, 323–341. [1347]
Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, Series B, 58, 267–288. [1358]
Tsui, K.-W., and Weerahandi, S. (1989), "Generalized *p*-Values in Significance Testing of Hypotheses in the Presence of Nuisance Parameters," *Journal of the American Statistical Association*, 84, 602–607. [1347]
——— (1991), "Corrections: "Generalized *p*-Values in Significance Testing of Hypotheses in the Presence of Nuisance Parameters,"*Journal of the American Statistical Association*, 86, 256 (*Journal of the American Statistical Association*, 84, 1989, 602–607). [1347]
Tukey, J. W. (1957), "Some Examples With Fiducial Relevance," *The Annals of Mathematical Statistics*, 28, 687–695. [1346]
Veronese, P., and Melilli, E. (2015), "Fiducial and Confidence Distributions for Real Exponential Families," *Scandinavian Journal of Statistics*, 42, 471–484. [1347]

Wandler, D. V., and Hannig, J. (2011), "Fiducial Inference on the Maximum Mean of a Multivariate Normal Distribution," *Journal of Multivariate Analysis*, 102, 87–104. [1347]

——— (2012a), "A Fiducial Approach to Multiple Comparisons," *Journal of Statistical Planning and Inference*, 142, 878–895. [1347]

——— (2012b), "Generalized Fiducial Confidence Intervals for Extremes," *Extremes*, 15, 67–87. [1347,1358]

Wang, J. C.-M., Hannig, J., and Iyer, H. K. (2012a), "Fiducial Prediction Intervals," *Journal of Statistical Planning and Inference*, 142, 1980–1990. [1347,1353]

——— (2012b), "Pivotal Methods in the Propagation of Distributions," *Metrologia*, 49, 382–389. [1347]

Wang, J. C.-M., and Iyer, H. K. (2005), "Propagation of Uncertainties in Measurements Using Generalized Inference," *Metrologia*, 42, 145–153. [1347]

——— (2006a), "A Generalized Confidence Interval for a Measurand in the Presence of Type-A and Type-B Uncertainties," *Measurement*, 39, 856–863. [1347]

——— (2006b), "Uncertainty Analysis for Vector Measurands Using Fiducial Inference," *Metrologia*, 43, 486–494. [1347]

Wang, Y. H. (2000), "Fiducial Intervals: What Are They?" *The American Statistician*, 54, 105–111. [1347]

Weerahandi, S. (1993), "Generalized Confidence Intervals," *Journal of the American Statistical Association*, 88, 899–905. [1347]

——— (1994), Correction: "Generalized Confidence Intervals," *Journal of the American Statistical Association*, 89, 726 (*Journal of the American Statistical Association*, 88, 1993, 899–905). [1347]

——— (1995), *Exact Statistical Methods for Data Analysis, Springer Series in Statistics,* New York: Springer-Verlag. [1347]

Welch, B. L., and Peers, H. W. (1963), "On Formulae for Confidence Points Based on Integrals of Weighted Likelihoods," *Journal of the Royal Statistical Society*, Series B, 25, 318–329. [1353]

Wilkinson, G. N. (1977), "On Resolving the Controversy in Statistical Inference," *Journal of the Royal Statistical Society*, Series B, 39, 119–171. [1346]

Xie, M., Liu, R. Y., Damaraju, C. V., and Olson, W. H. (2013), "Incorporating External Information in Analyses of Clinical Trials With Binary Outcomes," *The Annals of Applied Statistics*, 7, 342–368. [1347]

Xie, M., and Singh, K. (2013), "Confidence Distribution, the Frequentist Distribution Estimator of a Parameter: A Review," *International Statistical Review*, 81, 3–39. [1346,1348]

Xie, M., Singh, K., and Strawderman, W. E. (2011), "Confidence Distributions and a Unified Framework for Meta-Analysis," *Journal of the American Statistical Association*, 106, 320–333. [1346]

Xu, X., and Li, G. (2006), "Fiducial Inference in the Pivotal Family of Distributions," *Science in China*, Series A, 49, 410–432. [1347]

Yang, R., and Berger, J. O. (1997), "A Catalogue of Noninformative Priors," Technical Report ISDS 97–42, Duke University, Durham, NC. [1350]

Zhang, C., and Huang, J. (2008), "The Sparsity and Bias of the Lasso Selection in High-Dimensional Linear Regression," *The Annals of Statistics*, 36, 1567–1594. [1355]

Zhang, J., and Liu, C. (2011), "Dempster-Shafer Inference With Weak Beliefs," *Statistica Sinica*, 21, 475–494. [1346]